

gene  
CDS  
14072..14144,14960..15376)  
/gene="OSJNBa0014K08.2"  
join(8072..8086,9178..9387,9474..9671,12990..13120,  
14072..14144,14960..15376)  
/gene="OSJNBa0014K08.2"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAC05578.1"  
/db\_xref="GI:21902029"  
/translation="MRSFLVLPPLACGTGTPAPPPPTASLRRELALSPNLPARGG  
VAGQPMATREAPAAVSSARVYVPSLARRRRTPPPPPLFSPPVFSASP  
FGLIPMRSLRLRSGRSGDQAGALILSDVANYRGYLEDQCEQVEDQASAP  
NPFEEVYTFEKDSGRLSNIEOALCMFSEFOKLESLYFOLSCYTAEE  
OGIILADDLAKKPSROYMEKGSATNRAIDPFRSAAMSDAMRRARRSPCHANTR  
NGRRGARDPEPSKQPIRVLGRLSDDELWVTSAPAAVILRGETRAPAFALGTA  
GRRGEDDSVLR  
join(16191..16218,17138..17280)  
/gene="OSJNBa0014K08.3"  
join(16191..16218,17138..17280)  
/gene="OSJNBa0014K08.3"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAC05579.1"  
/db\_xref="GI:21902030"  
/translation="MAITVEMNQPPSVGHCCIMRSDVAATCRVQACSCITAPGCT  
GLREMAGLSSR"  
join(19222..19261,19694..19848,19915..20022,20628..20832,  
21203..21300,21397..21480,21572..21649,21807..21899,  
21891..22020,22277..22421,22872..22867,23073..23252,  
23347..23453,23552..23653,23956..24011,24094..24347,  
24422..24585,24909..25092,25170..25423)  
/gene="OSJNBa0014K08.4"  
join(19222..19261,19694..19848,19915..20022,20628..20832,  
21203..21300,21397..21480,21572..21649,21807..21899,  
21891..22020,22277..22421,22872..22867,23073..23252,  
23347..23453,23552..23653,23956..24011,24094..24347,  
24422..24585,24909..25092,25170..25423)  
/gene="OSJNBa0014K08.4"  
/note="contains ESTs AV065754(s10005),AU175140(s10005)"  
/codon\_start=1  
/product="putative kinase-like protein"  
/protein\_id="BAC05580.1"  
/db\_xref="GI:21902031"  
/translation="MGTNGEVEDPEDANRAVIMLGLLPEDLPDSDDELRLD  
YLINGELCYADKLMGVLEGWGGASDQRNVRKFLSVAAEMGLPFGVDLEEG  
SMSISIELCALNDVATQGHISSTAKPIRRKLELETDGPVLSAATPGRRPK  
SOORPLSGOKINEVVOFKGTITDLPARKISEMLNSLDNAPLOSILRVNGILD  
ESTERKGETPHRVHLIRVIOETIERGIDODHILRONSIIKTREDDYRSIKALE  
TVNGTNEEMAINRELVYVYKSKIDEKRGEGEDMILIREKMANIITSLIOE  
MQVNRNEQFRHDLKGLVDAENIRKVLNENKLFNEVELKGNIRVORARPEPLG  
QDOKLAIDYIGENGELLANPSKQKEGFMKFNKVFOTSSQAEVSDIOLPLRS  
VLGFNVICFAYGOGTGKTYTSGPOTSEDEGVNRYALNDLFLISLKNAPSES  
NRAVGTALNERSRSHSILTVHVRGLDVYNSGTSRGLILDLAGSEREREADAG  
RLKEACHINKSLGIDVILSLAOKNAHVYRSKLTLOVYOSGLGQAKIMLEVINP  
DISEYFTITLKFAERVSGVELGAARNSGEDIKELLFOVANSIDITARKMETEO  
LOILKSPKSMKTDNGSNLROSTISSTGSLSPVASONQOISVYITPLPLATQGLAG  
SVEAEADNADSDGCVGETEYSPAGASETSABRAAYLLIFSTISCAKSCACDTGSE  
LMTVPRLRGT"  
join(27978..28537,28795..29152,29378..29650)  
/gene="OSJNBa0014K08.5"  
join(27978..28537,28795..29152,29378..29650)  
/gene="OSJNBa0014K08.5"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAC05581.1"  
/db\_xref="GI:21902032"  
/translation="MATASVYTLQPRELPPVEERAPPPSPGVYLLSSGSLAG  
ECPDNDDEIGRLRRSARVPLRLPERAKKAKAMPVIVYLLALPAGGPV  
AEALRSAVAFGCGVYGVHGVDSIVSALRLHAAALARAATPEPEVEVNGDDENE

gene  
CDS  
35132..35437)  
/gene="OSJNBa0014K08.6"  
complement(join(31081..31442,31391..31579,31787..31815,  
31903..31993,32132..32306,32410..32549,32638..32770,  
33508..33574,34086..34243,34332..34405,34496..34609,  
35132..35437))  
/gene="OSJNBa0014K08.6"  
/note="contains ESTs C79349(E2532),C72949(E2532)"  
/codon\_start=1  
/product="putative casein kinase-like protein"  
/protein\_id="BAC05582.1"  
/db\_xref="GI:21902033"  
/translation="MPVLSARARRAREQENPAAAAAPPPARRRRARRREPEY  
AVEAPEAEEREEIEVADLREGGGKREKEDSGARSADKQAVDEGVTPEPT  
VOVGSSPYRVADKIKGSGQGVYVVRMLNMGASEVALKEHRTSGCNHCPPE  
MOVYNAIGIIGVPRVYHKKGRGEYVYVMTPHDYERNCLFALSKILINLYERS  
TKWRASGLGHVDQDPDPREVVEYVYVNTKDEEPYAKCISLEFTRVNPDIRP  
INTDCAOKLILHVGOKRGRLMEETDEOPKKIKMGDPATOWISVYARPMKORAY  
NVADSLVOHIDKNGEDSLFTSCITSCSNMALLINDAGTGTSTVYELSPLEKMT  
MDWERNYITILAGNANGSSLVMSKOTMTQOSKYVELDELFLPSGHIQRNDNG  
RITAPAAATWQDAFVLSVRRKPIDETQETLRTSAFSPQHVKKSGHLEVALSSRLA  
LVSQSAAYVYMLASSECCLAYPDRYAVVYKFNQYRS"  
join(39834..39907,40580..40722,41299..41408)  
/gene="OSJNBa0014K08.7"  
join(39834..39907,40580..40722,41299..41408)  
/gene="OSJNBa0014K08.7"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAC05583.1"  
/db\_xref="GI:21902034"  
/translation="MPYVGTGNSCAEYIRSLISROKEEYKPGKSELNPM  
AVGSLDAKSCFKHICDYSERLCVSIENRFARVSLDVQNRKRMVYVLOCLOI  
SLFFPC"  
complement(join(47169..47198,47300..47377,47591..47689,  
47772..47834,48987..49058,49345..49219,49310..49366,  
50328..50531,50662..50739,50830..50892,50991..51629,  
51712..51761,51841..51921,52241..52268))  
/gene="OSJNBa0014K08.8"  
complement(join(47169..47198,47300..47377,47591..47689,  
47772..47834,48987..49058,49345..49219,49310..49366,  
50328..50531,50662..50739,50830..50892,50991..51629,  
51712..51761,51841..51921,52241..52268))  
/gene="OSJNBa0014K08.8"  
/note="hypothetical protein  
similar to Arabidopsis thaliana chromosome 2, At2g19950"  
/codon\_start=1  
/protein\_id="BAC05584.1"  
/db\_xref="GI:21902035"  
Query Match 81.6%; Score 20.4; DB 8; Length 177263;  
Best Local Similarity 95.5%; Pred. No. 2,1e+02;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3 ATATATCTACATCAACA 24  
|||||  
Db 92833 ATATATCTACATCAACA 92812  
|||||  
RESULT 10  
AC115717 226726 bp DNA linear HTG 26-JUN-2002  
LOCUS  
DEFINITION Mus musculus clone RP23-28P10, WORKING DRAFT SEQUENCE, 9 ordered  
pieces.  
AC115717  
AC115717 GI:21592233  
VERSION

```

misc_feature      5204. .5675
                  /note="match: GSS: Em:AQ712434"
repeat_region    5702. .5976
                  /note="LRR16A repeat: matches 132. .428 of consensus
misc_feature      6212. .6878
                  /note="match: GSS: Em:AQ321209"
misc_feature      6212. .6775
                  /note="match: GSS: Em:AQ377996"
misc_feature      6213. .6543
                  /note="match: GSS: Em:AQ382827"
misc_feature      6216. .6770
                  /note="match: GSS: Em:AQ344327"
misc_feature      6221. .6541
                  /note="match: GSS: Em:AQ052109"
misc_feature      6229. .6703
                  /note="match: GSS: Em:AQ370102"
misc_feature      6254. .6541
                  /note="match: GSS: Em:AQ214603"
misc_feature      6256. .6733
                  /note="match: GSS: Em:AQ454556"
misc_feature      6272. .6541
                  /note="match: GSS: Em:AQ095672"
misc_feature      6329. .6686
                  /note="match: GSS: Em:AQ629712"
misc_feature      6397. .6873
                  /note="match: GSS: Em:AQ587044"
misc_feature      6406. .6787
                  /note="match: GSS: Em:AQ587048"
misc_feature      6546. .6898
                  /note="match: GSS: Em:AQ052109"
misc_feature      6546. .6741
                  /note="match: GSS: Em:AQ095672"
misc_feature      6546. .6739
                  /note="match: GSS: Em:AQ214603"
repeat_region    7281. .7429
                  /note="MER65A repeat: matches 295. .445 of consensus
                  MER65A repeat: matches 295. .445 of consensus"
repeat_region    7454. .7774
                  /note="MER65A repeat: matches 15. .334 of consensus
                  MER65A repeat: matches 15. .334 of consensus"
repeat_region    8040. .8082
                  /note="MDEL repeat: matches 34. .77 of consensus
                  MDEL repeat: matches 34. .77 of consensus"
repeat_region    8497. .8643
                  /note="LIPB1 repeat: matches 6008. .6155 of consensus
                  LIPB1 repeat: matches 6008. .6155 of consensus"
repeat_region    8933. .9240
                  /note="LRR40a repeat: matches 2. .327 of consensus
                  LRR40a repeat: matches 2. .327 of consensus"
repeat_region    9246. .10036
                  /note="L1 repeat: matches 4060. .4853 of consensus
                  L1 repeat: matches 4060. .4853 of consensus"
repeat_region    10042. .10325
                  /note="LRR14A repeat: matches 1. .344 of consensus
                  LRR14A repeat: matches 1. .344 of consensus"
repeat_region    10326. .12101
                  /note="HERVK4 repeat: matches 1. .1888 of consensus
                  HERVK4 repeat: matches 1. .1888 of consensus"
repeat_region    12102. .12546
                  /note="MER9 repeat: matches 1. .509 of consensus
                  MER9 repeat: matches 1. .509 of consensus"
repeat_region    12549. .15363

```

Every Match	81.6%;	Score 20.4;	DB 9;	Length 127682;
Local Similarity	95.5%;	Pred. No. 2.3e+02;		
Matches 21;	Conservative	0;	Mismatches 1;	Indels 0;
				Gaps 0

QY	4	TATATATCTACATTTCAAAACAA	25
Db	60289	TGTATATCTACATTTCAAAACAA	60268

RESULT 9  
AP003376/c  
LOCUS 177263 bp DNA linear PLN 18-JUL-2002  
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,  
BAC clone:OSUNBa0014K08.  
ACCESSION AP003376  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,  
clone:OSUNBa0014K08.  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Echinozoidae; Oryzaceae; Oryza.  
REFERENCE  
1  
Sasaki,T., Matsumoto,T. and Yamamoto,K.  
Oryza sativa nipponbare(GAS3) genomic DNA, chromosome 1, BAC  
clone:OSUNBa0014K08  
JOURNAL Published Only in Database (2001)  
REFERENCE 2 (bases 1 to 177263)  
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.  
TITLE Direct Submission  
JOURNAL Submitted (07-MAR-2001) Takuji Sasaki, National Institute of  
Agrobiological Sciences, Rice Genome Research Program, Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail:tasakia@affrc.go.jp, URL:http://rpg.dna.affrc.go.jp/,  
Tel:81-298-38-7441, Fax:81-298-38-7408)  
COMMENT On Jul 17, 2002 this sequence version replaced gi:20975475.  
The previous version of this record is the following:

On Jul 17, 2002, this sequence version replaced g12073622. Genes were predicted from the integrated results of the following: GENSCAN1.0, BLASTX.0, BLASTZ.0 as well as SplicePredictor (October 1998 version). The genomic sequence was searched against NCBI NonRedundant Protein database, nr (<http://ncbi.nlm.nih.gov/blast/cd>) and the cDNA sequence database at RGP. Protein homologs of the coding regions were searched againsts NCBI NonRedundant Protein database with BLASTX.0. ESTs represent the identified cDNA sequences using BLASTN 2.0 with the corresponding DDBJ accession no. and RGP clone ID. A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level, such as same name, 'putative' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein.

The orientation of the sequence is from M13rev to -21M13 of the BAC clone. This sequence of OSUNA0014K08 clone has an overlap with P0433E07 clone (DDBJ : AP003168) at the position 1 to 45,852 of 5' end and an overlap with P0481E12 clone (DDBJ : AP003076) at the position 172,063 to 177,263 of 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rgp.cna.affrc.go.jp/genomeseq.html>.

Source

1. /r17203  
/organism="Oryza sativa (japonica cultivar-group)"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/chromosome="1"  
/clone="OSJNBa0014K08"  
complement(join((1396..1701,1758..2043,2410..2544,  
3405..3547,5098..5199,6055..6167,6605..6690,7352..7491))  
/gene="OSJNBa0014K08.1"  
complement(join(1396..1701,1758..2043,2410..2544,  
3405..3547,5098..5199,6055..6167,6605..6690,7352..7491))  
/gene="OSJNBa0014K08.1"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAC05577.1"  
/db\_xref="gi:21902028"  
/translation="MAEPDILEREQVEQIILDEMLQEEVDAGSSSEVDI  
ELRNTHDGGSRSEALFNTVSFLPTDGGVLEPEELIRVQPSRLTAVDKAV  
HYVADPCMIGVTHHTKOLDGSSNVTYQGNFRRRLRIADAVALLSRSDGVGA  
VYKQSCGGEDVTYNATGLRGAPSKAHSLFEPPTALPQRGEKRCRHGGHID  
AIAHLKRGFLELAKKRGNSPQRKGRARLRVLPVGNEDDPPPLAPALFLGR



Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CCATATATATCATTCATAAA 22  
 |||  
 Db 38714 CCTATATATCATTCATAAA 38735

RESULT 8  
 AL137861/c 127682 bp DNA linear PRI 01-NOV-2000  
 LOCUS Human DNA sequence from clone RP4-813B7 on chromosome 1 contains a  
 DEFINITION pseudogene similar to zinc finger protein ZNF186, STS and GSSS,  
 complete sequence.

ACCESSION AL137861  
 VERSION AL137861.5 GI:9187172  
 KEYWORDS HTG.

SOURCE human

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 127682)  
 AUTHORS Cobley, V.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk

COMMENT

On Jul 14, 2000 this sequence version replaced gi:8694204.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence has been finished according to sequence map criteria  
 as follows. An attempt is made to resolve all sequencing problems,  
 such as compressions and repeats, but not necessarily within known  
 annotated human repeat sequence elements (e.g. Alu). Where the  
 sequence is ambiguous, there is an annotation using the 'unsure'  
 feature key.

The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em, EMBL; SW, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information  
 on the WORMPEP database can be found at:  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human  
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
 Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr1

RP4-813B7 is from the library RPCI-4 constructed at the Roswell  
 Park Cancer Institute by the group of Pieter de Jong. For further  
 details see http://bacpac.med.buffalo.edu/  
 VECTOR: pCYPAC2

This sequence is the entire insert of clone RP4-813B7.

FEATURES  
 source

1..127682  
 location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /clone="RP4-813B7"  
 /clone\_1fb="RPCI-4"  
 1..235  
 repeat\_region  
 /note="AluY repeat: matches 75, .310 of consensus  
 AluY repeat: matches 75, .310 of consensus"  
 239..538  
 repeat\_region  
 /note="L1M4 repeat: matches 2107, .2438 of consensus"  
 336..538  
 repeat\_region  
 /note="L1M4 repeat: matches 2227, .2438 of consensus"  
 550..1553  
 repeat\_region  
 /note="L1M4 repeat: matches 2582, .3669 of consensus"  
 complement(565..1283)  
 misc\_feature  
 /note="match: GSS: Em:AQ383567"  
 675..1155  
 repeat\_region  
 /note="L1M4 repeat: matches 2732, .3226 of consensus"

misc\_feature complement(698..1442)  
 /note="match: GSS: Em:AQ896860"  
 misc\_feature complement(906..1442)  
 /note="match: GSS: Em:AQ761517"  
 1288..1539  
 misc\_feature  
 /note="match: GSS: Em:AQ879547"  
 1290..1693  
 misc\_feature  
 /note="match: GSS: Em:AQ826107"  
 1294..2055  
 misc\_feature  
 /note="match: GSS: Em:AQ301298"  
 1294..1860  
 misc\_feature  
 /note="match: GSS: Em:AQ471100"  
 1362..2055  
 misc\_feature  
 /note="match: GSS: Em:AQ490007"  
 1362..1961  
 misc\_feature  
 /note="match: GSS: Em:AQ585548"  
 1362..1761  
 misc\_feature  
 /note="match: GSS: Em:AQ275646"  
 1447..1837  
 misc\_feature  
 /note="match: GSS: Em:AQ355624"  
 1448..1964  
 misc\_feature  
 /note="match: GSS: Em:AQ474420"  
 1449..2054  
 misc\_feature  
 /note="match: GSS: Em:B76221"  
 1450..2017  
 misc\_feature  
 /note="match: GSS: Em:AQ537460"  
 1451..2223  
 misc\_feature  
 /note="match: GSS: Em:AQ738677"  
 1451..2150  
 misc\_feature  
 /note="match: GSS: Em:AQ739952"  
 1451..1896  
 misc\_feature  
 /note="match: GSS: Em:AQ669827"  
 1451..1877  
 misc\_feature  
 /note="match: GSS: Em:AQ810603"  
 1451..1827  
 misc\_feature  
 /note="match: GSS: Em:AQ718640"  
 1466..1877  
 misc\_feature  
 /note="match: GSS: Em:AQ692785"  
 1592..2208  
 repeat\_region  
 /note="L1MB1 repeat: matches 5338, .5979 of consensus  
 L1MB1 repeat: matches 5338, .5979 of consensus"  
 2199..2404  
 repeat\_region  
 /note="L1M4 repeat: matches 3935, .4135 of consensus  
 L1M4 repeat: matches 3935, .4135 of consensus"  
 complement(2626..3066)  
 misc\_feature  
 /note="match: GSS: Em:AQ245895"  
 complement(2887..3061)  
 misc\_feature  
 /note="match: GSS: Em:B91023"  
 3067..3302  
 misc\_feature  
 /note="match: GSS: Em:AQ009565"  
 3074..3453  
 misc\_feature  
 /note="match: GSS: Em:AQ261089"  
 3074..3302  
 misc\_feature  
 /note="match: GSS: Em:AQ059372"  
 3407..3448  
 repeat\_region  
 /note="XMR5 repeat: matches 1, .42 of consensus"  
 complement(3620..4114)  
 misc\_feature  
 /note="match: GSS: Em:AQ117010"  
 complement(3956..4111)  
 misc\_feature  
 /note="match: GSS: Em:AQ304953"  
 4136..4803  
 misc\_feature  
 /note="match: GSS: Em:AQ531955"  
 4636..4942  
 repeat\_region  
 /note="AluSc repeat: matches 1, .308 of consensus  
 AluSc repeat: matches 1, .311 of consensus"  
 5143..5891  
 misc\_feature  
 /note="match: GSS: Em:AQ224449"  
 5147..5738  
 misc\_feature  
 /note="match: GSS: Em:B60185"  
 5147..5659  
 misc\_feature  
 /note="match: GSS: Em:AQ076436"  
 5154..5605  
 misc\_feature  
 /note="match: GSS: Em:AQ608022"

/db\_xref="taxon:4896"  
/chromosome="11"  
/map="11R"  
/clone="p1 p857"  
complement(join(1572..1631,1701..1836,1885..1997))  
/gene="SPB8B7.01c"  
complement(join(1572..1631,1701..1836,1885..1997))  
/gene="SPB8B7.01c"  
/note="SPB8B7.01c, len:102"  
/codon\_start=1  
/label="SPB8B7.01c  
/product="hypothetical protein"  
/protein\_id="CAA21786.1"  
/db\_xref="GI:3810825"  
/db\_xref="SPTREMBL:O94251"  
/translation="MSVERIRRAAPAFKGIIRKKTLLIAMIRMRKKLDIDSLIO  
ATGMAIPHLVELSDKRYKRGQINISMDTATVEDEIVDEVSIONKRLGNVLLRVR  
"  
misc\_feature  
complement(1632..1646)  
/gene="SPB8B7.01c"  
/note="ctaccagctgtag, splice branch and acceptor"  
complement(1655..1700)  
/gene="SPB8B7.01c"  
/note="gtacgc, splice donor sequence"  
complement(1837..1852)  
/gene="SPB8B7.01c"  
/note="ctaccagagcttag, splice branch and acceptor"  
complement(1879..1884)  
/gene="SPB8B7.01c"  
/note="gtacgt, splice donor sequence"  
/gene="SPB8B7.02"  
2917..3702  
/gene="SPB8B7.02"  
/note="SPB8B7.02, len:261, SIMILARITY:Oxytricha fallax,  
0.0905, hypothetical 129.9 kd protein., (1088 aa), fasta  
seq res: opt: 119, E():3.4e-05, (21.3% identity in 254  
aa)"  
/codon\_start=1  
/label="SPB8B7.02  
/product="hypothetical protein."  
/protein\_id="CAA21787.1"  
/db\_xref="GI:3810826"  
/db\_xref="SPTREMBL:O94252"  
/translation="MNDESHISVLPVNNKQTSVLQNLKEETENOLKEKKRSQHL  
KEEINLQLVAINDLRSKTEKELKANKEDTPTNSFASGDLTANKKIQRSQRL  
EESLISYKKVKYEMKORHSASMKLELAKLSAOLQTNVNLRSNNDEILQDISER  
KDCRELIAEQLQASLINDRONSDRIAKNHSSTIDPSRAHRSYLNATKNYTHID  
SLTTRHDSQEDVSVEKLNRSKISELEKLPADYCSIEDAFD"  
gene  
complement(4027..5118)  
/gene="SPB8B7.03c"  
complement(4027..5118)  
/gene="SPB8B7.03c"  
/note="SPB8B7.03c, len:363"  
/codon\_start=1  
/label="SPB8B7.03c  
/product="60s ribosomal protein 12"  
/protein\_id="CAA21788.1"  
/db\_xref="GI:3810827"  
/db\_xref="SPTREMBL:O94253"  
/translation="MAARPTVSYKSGVSSSELTIALPVEFKAPIREDIYRVHTAY  
AKNRQPTAVSEKAGHOTSAESWGTGRLLATPRYGGGTTRSGOAAFGNCRGRMT  
APKTWRWAVKVNQNERVYAISSAVASGVPSLLARGHRIETIPEVLVDDAVOS  
FOKTEAVALLKEIKAYRDVYKVANSRKLKRGKLRNRRHVORGRPLVNEAGIV  
KAFRNIPGEIVNRRNLQIAPGSHLGRFVITWKSAFGLDVSGSTPAQIKKN  
YFLPNISNADVTRLINSDEIOSVKAAGSRVYKRAHVQKKNPLKNAVALRLNPPYA  
KAYKANVLTNGKTPKAGSEKFTVLEHN"  
misc\_feature  
complement(4330..5064)  
/gene="SPB8B7.03c"  
/note="Pfam match to entry PF00573 L1e, Ribosomal protein  
family 11e"  
join(5912..5983,6026..8413)  
/gene="SPB8B7.04"  
gene

CDS  
join(5912..5983,6026..8413)  
/gene="SPB8B7.04"  
/note="SPB8B7.04, len:819"  
/codon\_start=1  
/label="SPB8B7.04  
/product="hypothetical protein"  
/protein\_id="CAA21789.1"  
/db\_xref="GI:3810828"  
/db\_xref="SPTREMBL:O94254"  
/translation="MKNLNIRKYNNDNELLADMLILKLSRSSDCLDLRBRKW  
ELVCKEELINGERNPLNPLPELPVNSLELQSYNDMPFKHYSONVNLNKTQYDF  
TEQLIRKSLNKLNGTLRETRSPSFSISLSEISNALDSCFNDMTLISESP  
IDGMLLEAMDSONDLFTSKOLNNGPNPEVETFELESNMGLSQRAATSTTPVCN  
ISSVTAINSGVEISNASHSSSTSELPCTGNTSSIFQVNMENSKAISBGYIADH  
OQKRVSONIKKEFLPEEFSELEGLVNNKFSMLYVLSETEKTRDSKYFTFRL  
VTKTRPLVASESFVEETKNEFTPEPLCAINPHIDSCDMDRLPSLKSSVAAPID  
KIKREGSRKSFOSLIFQPNHCIFRFSNINYSKOSQNLITLTTIRPICAAYOKLEY  
NITVLNBERKSSGSLILSASPTCTENTYENTRHRFSKQIPEDTVISFDVOCV  
VTEATRTLELSTLSESLYETALKLPVIRYVGDNRFCFQLEPYDIYLRKLLAD  
WRLLTNVLIQELISDYCHEKLRKVRNLIQITNPQOEFLHRLPTIRILQISREY  
FLHSSSYIFNDAREDDVGVLSKMMWISMAVDCANFCQKDGTFPRTSRKQI  
LILEGILIKPERYNLFKRYVLPQPLPDSRTLYEPOTIYVNTSLTFDGSFYSO  
NGGDPPEPRONTYTPDKSFELLQNSYISYGAKKIINKLLELIVLVNVTAM  
CLTLLIGITLILQVVFITN"  
5984..5989  
/gene="SPB8B7.04"  
/note="gtaatt, splice donor sequence"  
6013..6025  
/gene="SPB8B7.04"  
/note="ctaccagctgtag, splice branch and acceptor"  
complement(8688..9422)  
/gene="SPB8B7.05c"  
complement(8688..9422)  
/gene="SPB8B7.05c"  
/note="SPB8B7.05c, len:244, SIMILARITY:porphyrinidum  
purpureum., Q43060, carbonic anhydrase, (571 aa), fasta  
scores: opt: 580, E():2.2e-31, (41.68 identity in 221 aa)"  
/codon\_start=1  
/label="SPB8B7.05c  
/product="carbonic anhydrase"  
/protein\_id="CAA21790.1"  
/db\_xref="GI:3810829"  
/db\_xref="SPTREMBL:O94255"  
/translation="MDRLKRRILEIDHDEITIDREASTSPVSGAKIDQNEIIDL  
LERNLISQOTSRRKPYEFATKDIDQPPQVLMGSDSRVPEFTILNLNLEGVFRN  
IANVVPSSDINALAVMEYSVTYKVKHIIYCGVYGGVAAAGPMLNMLDMLHRI  
RDVEDRRELDALDEPOLRLKLAELNTRQAISVTRGVFRAMEKRGLOVHGWIY  
DLSNGQIKKIDITDAIKKAKYGYDS"  
misc\_feature  
complement(8754..8822)  
/gene="SPB8B7.05c"  
/note="Pfam match to entry Pf00484 Pro\_CA,  
prokaryotic-type carbonic anhydrases"  
complement(8937..9272)  
/gene="SPB8B7.05c"  
/note="Pfam match to entry PF00484 Pro\_CA,  
prokaryotic-type carbonic anhydrases"  
join(10661..10969,11033..11056)  
/gene="SPB8B7.06; rpl4-2"  
join(10661..10969,11033..11056)  
/note="SPB8B7.06; rpl4-2"  
/codon\_start=1  
/label="rpl4  
/product="60s acidic ribosomal protein L4.2/L48"  
/protein\_id="CAA21791.1"  
/db\_xref="GI:3810830"  
/db\_xref="SPTREMBL:O94256"  
/translation="MKYLAAYLITLVGSGKSPSADIESVLTGIGAESERIETLIN  
ELNKGIDDELIAENKRLATVPTGGASAPAPAAAGAAAPAAEAAAEKAEKKEDESE  
DMGRTLD"  
Query Match 81.6%; Score 20.4; DB 8; Length 71000;  
Best Local Similarity 95.5%; Pred. No. 2.6e+02;

```

*      78100      80648: contig of 2549 bp in length
*      80649      80748: gap of unknown length
*      80749      83618: contig of 2870 bp in length
*      83619      83718: gap of unknown length
*      83719      86486: contig of 2768 bp in length
*      86487      86586: gap of unknown length
*      86587      89209: contig of 2623 bp in length
*      89210      89309: gap of unknown length
*      89310      91972: contig of 2663 bp in length
*      91973      92072: gap of unknown length
*      92073      94748: contig of 2676 bp in length
*      94749      94848: gap of unknown length
*      94849      97725: contig of 2878 bp in length

```

```

Query Match      83.2%; Score 20.8; DB 2; Length 152557;
Best Local Similarity 91.7%; Pred. No. 1.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2 CATATATATCTACTTCAAAACAA 25
Db      87874 CATATATATATCATGCAAAACAA 87897

```

```

RESULT 6
AC092723      188458 bp      DNA      linear      PRI 06-FEB-2002
LOCUS      Homo sapiens chromosome 16 clone RP11-542M13, complete sequence.
DEFINITION      AC092723 AC068615
ACCESSION      AC092723.3 GI:18542980
VERSION      HTG.
KEYWORDS
SOURCE      Homo sapiens.
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS      1 (bases 1 to 188458)
TITLE      DOE Joint Genome Institute.
JOURNAL      Sequencing of Human Chromosome 16
REFERENCE
AUTHORS      2 (bases 1 to 188458)
TITLE      DOE Joint Genome Institute.
JOURNAL      Direct Submission
AUTHORS      Submitted (21-JUL-2001) Production Sequencing Facility, DOE Joint
      Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
      3 (bases 1 to 188458)
      DOE Joint Genome Institute.
      Direct Submission
      Submitted (06-FEB-2002) Production Sequencing Facility, DOE Joint
      Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
      On Feb 6, 2002 this sequence version replaced g1:15055296.
COMMENT
      Sequence Quality Assessment:
      This entry has been annotated with sequence quality
      estimates computed by the Phrap assembly program.
      All manually edited bases have been reduced to quality zero.
      Quality levels above 40 are expected to have less than
      1 error in 10,000 bp.
      Base-by-base quality values are not generally visible from the
      GenBank flat file format but are available as part
      of this entry's ASN.1 file.

```

```

FEATURES
Source      Location/Qualifiers
            1..188458
            /organism="Homo sapiens"
            /db_xref="taxon:9606"

```

```

BASE COUNT      43597 a 48044 c 45725 g 45092 t
ORIGIN

```

```

Query Match      83.2%; Score 20.8; DB 9; Length 188458;
Best Local Similarity 91.7%; Pred No. 1.5e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2 CATATATATCTACTTCAAAACAA 25
Db      65620 CATCTATATCTACTTCAAAACAA 65643

```

```

RESULT 7
SPB8B7
LOCUS      S. pombe chromosome II pl p8B7.
DEFINITION      AL032684
ACCESSION      AL032684.1 GI:3810824
VERSION
KEYWORDS      60S ribosomal protein L4.2/L4B; autophagy; carbonic anhydrase; cell
      division control; transcriptional activator; cyclin; dbp2; DNA
      polymerase epsilon subunit b; ER lumen protein retaining receptor;
      LTR; NOL1 family; NOP2 family; p68-like protein; para-aminobenzoate
      synthase; peptidyl-prolyl cis-trans isomerase; puc1; RNA-binding;
      rpa2; rpl4-2; sun family; thiamine biosynthesis; tRNA nuclear
      export; ubiquitin carboxyl-terminal hydrolase; ubiquitin-protein
      ligase; zinc-finger.
      Schizosaccharomyces pombe.
      Schizosaccharomyces pombe.
      Schizosaccharomycetes pombe
      Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
      Schizosaccharomycetales; Schizosaccharomycetaceae;
      Schizosaccharomyces.

```

```

SOURCE
ORGANISM
REFERENCE
AUTHORS      1 (bases 1 to 71000)
      Beck, A., Reinhardt, R., Lyne, M., Rajandream, M.A. and Barrell, B.G.
TITLE      Direct Submission
JOURNAL      Submitted (29-OCT-1998) European Schizosaccharomyces genome
      sequencing project, Sanger Centre, The Wellcome Trust Genome
      Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
      and Max Planck Institut fuer Molekulare Genetik, Innestr. 73,
      D-14195 Berlin, Germany

```

## COMMENT

Notes: Details of yeast sequencing at the Sanger Centre are available on the World Wide Web.  
 (URL, <http://www.sanger.ac.uk/Projects/S.pombe/>)  
 During 1995 to 1996 about 66% of S. pombe chromosome I was sequenced by the Sanger Centre. The sequencing from The European Commission is now being continued with funding from The European Commission. Fourteen European sequencing laboratories, including the Sanger Centre, are participating in the project.  
 Protein coding regions (CDS) have been predicted with the help of computer analysis using the GeneFinder program in Pombase (an ACEDB database) with additional predictions for the branch-acceptor sites supplied by the program Splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.  
 CDS are numbered using the following system eg SPBC25H2.01c. SP (S. pombe). B (chromosome 2). c25H2 (cosmid name). .01 (first CDS). c (complementary strand).  
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
 The length in codons is given for each CDS.  
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
 pl p8B7 is overlapped at the 3' end by cosmid cl9F5. EMBL entry SPBC19F5, accession number AL022599.

```

FEATURES
Source      Location/Qualifiers
            1..71000
            /organism="Schizosaccharomyces pombe"
            /strain="972h-"

```

Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,  
 Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
 Scherer, S., Scott, G., Shen, H., Shooshani, N., Sison, I.,  
 Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Sytek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmami, K., Vaequez, L., Vera, V., Villalob, D., Vinsom, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleciyk, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y., Zhou, J., Zortilla, S., Nelson, D.,  
 Weinstein, G. and Gibbs, R.

Unpublished  
 2 (bases 1 to 152557)  
 Worley, K.C.

Direct Submission  
 Submitted (09-JUN-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 152557)  
 Worley, K.C.

Direct Submission  
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 12, 2002 this sequence version replaced gi:18092762.

----- Genome Center -----  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 Project Information  
 Center project name: CH230-232018  
 Sequencing Statistics  
 Chemistry: Dye-terminator Big Dye 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 99284 bases at least Q40  
 Consensus quality: 104691 bases at least Q30  
 Consensus quality: 109276 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 70 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

15106 15205: gap of unknown length  
 15206 16312: contig of 1107 bp in length  
 15207 16413: gap of unknown length  
 15208 16413: contig of 1642 bp in length  
 15209 18054: gap of unknown length  
 15210 18054: gap of unknown length  
 15211 18155: contig of 1851 bp in length  
 15212 20005: gap of unknown length  
 15213 20105: gap of unknown length  
 15214 20106: contig of 1261 bp in length  
 15215 21367: gap of unknown length  
 15216 21367: contig of 1176 bp in length  
 15217 22643: gap of unknown length  
 15218 22643: gap of unknown length  
 15219 22743: contig of 1517 bp in length  
 15220 24258: gap of unknown length  
 15221 24258: gap of unknown length  
 15222 24360: contig of 2133 bp in length  
 15223 26493: gap of unknown length  
 15224 26592: gap of unknown length  
 15225 28144: contig of 1552 bp in length  
 15226 28244: gap of unknown length  
 15227 28245: gap of unknown length  
 15228 28245: contig of 1415 bp in length  
 15229 29661: gap of unknown length  
 15230 29661: gap of unknown length  
 15231 29761: gap of unknown length  
 15232 31395: contig of 1635 bp in length  
 15233 31395: gap of unknown length  
 15234 31495: gap of unknown length  
 15235 31495: contig of 1666 bp in length  
 15236 33161: contig of 1666 bp in length  
 15237 33161: gap of unknown length  
 15238 33262: contig of 1135 bp in length  
 15239 34396: gap of unknown length  
 15240 34396: gap of unknown length  
 15241 34497: gap of unknown length  
 15242 36082: contig of 1586 bp in length  
 15243 36082: gap of unknown length  
 15244 36183: gap of unknown length  
 15245 37418: contig of 1236 bp in length  
 15246 37518: gap of unknown length  
 15247 37519: gap of unknown length  
 15248 39501: contig of 1982 bp in length  
 15249 39501: gap of unknown length  
 15250 39600: gap of unknown length  
 15251 41458: contig of 1858 bp in length  
 15252 41458: gap of unknown length  
 15253 41559: gap of unknown length  
 15254 42950: contig of 1392 bp in length  
 15255 42951: gap of unknown length  
 15256 43050: gap of unknown length  
 15257 44383: contig of 1333 bp in length  
 15258 44383: gap of unknown length  
 15259 44384: gap of unknown length  
 15260 46166: contig of 1582 bp in length  
 15261 46166: gap of unknown length  
 15262 47415: contig of 1150 bp in length  
 15263 47415: gap of unknown length  
 15264 47515: gap of unknown length  
 15265 49478: contig of 1963 bp in length  
 15266 49478: gap of unknown length  
 15267 49579: gap of unknown length  
 15268 51179: contig of 1601 bp in length  
 15269 51180: gap of unknown length  
 15270 51280: gap of unknown length  
 15271 52639: contig of 1360 bp in length  
 15272 52639: gap of unknown length  
 15273 52740: gap of unknown length  
 15274 54083: contig of 1344 bp in length  
 15275 54084: gap of unknown length  
 15276 54184: gap of unknown length  
 15277 56391: contig of 2208 bp in length  
 15278 56392: gap of unknown length  
 15279 56491: gap of unknown length  
 15280 56492: gap of unknown length  
 15281 57617: contig of 1125 bp in length  
 15282 57617: gap of unknown length  
 15283 58769: contig of 1053 bp in length  
 15284 58769: gap of unknown length  
 15285 58869: gap of unknown length  
 15286 60433: contig of 1564 bp in length  
 15287 60433: gap of unknown length  
 15288 60534: gap of unknown length  
 15289 60534: contig of 1585 bp in length  
 15290 62118: gap of unknown length  
 15291 62118: gap of unknown length  
 15292 62218: gap of unknown length  
 15293 64323: contig of 2105 bp in length  
 15294 64323: gap of unknown length  
 15295 64424: gap of unknown length  
 15296 64424: contig of 2344 bp in length  
 15297 66767: gap of unknown length  
 15298 66767: gap of unknown length  
 15299 66867: gap of unknown length  
 15300 66867: contig of 2096 bp in length  
 15301 68863: gap of unknown length  
 15302 68863: gap of unknown length  
 15303 68864: gap of unknown length  
 15304 68864: contig of 2779 bp in length  
 15305 71842: gap of unknown length  
 15306 71842: gap of unknown length  
 15307 73906: contig of 1964 bp in length  
 15308 73906: gap of unknown length  
 15309 74007: gap of unknown length  
 15310 74007: contig of 2437 bp in length  
 15311 76443: gap of unknown length  
 15312 76443: gap of unknown length  
 15313 77939: contig of 1456 bp in length  
 15314 78000: gap of unknown length

*	1	1360:	contig of 1360 bp in length
*	1361	1460:	gap of unknown length
*	1461	2510:	contig of 1050 bp in length
*	2511	2610:	gap of unknown length
*	2611	3721:	contig of 1111 bp in length
*	3722	3821:	gap of unknown length
*	3822	4930:	contig of 1109 bp in length
*	4931	5030:	gap of unknown length
*	5031	6334:	contig of 1304 bp in length
*	6335	6434:	gap of unknown length
*	6435	7522:	contig of 1068 bp in length
*	7523	7622:	gap of unknown length
*	7623	8669:	contig of 1247 bp in length
*	8670	8969:	gap of unknown length
*	8970	10303:	contig of 1334 bp in length
*	10304	10403:	gap of unknown length
*	10404	11605:	contig of 1202 bp in length
*	11606	11705:	gap of unknown length
*	11706	12960:	contig of 1275 bp in length
*	12961	13080:	gap of unknown length
*	13081	14163:	contig of 1083 bp in length
*	14164	14263:	gap of unknown length
*	14264	15815:	contig of 1552 bp in length
*	15816	15915:	gap of unknown length
*	15916	17741:	contig of 1826 bp in length
*	17742	17841:	gap of unknown length
*	17842	19575:	contig of 1734 bp in length
*	19576	19675:	gap of unknown length
*	19676	20867:	contig of 1132 bp in length
*	20868	20967:	gap of unknown length
*	20969	22586:	contig of 1619 bp in length
*	22587	22686:	gap of unknown length
*	22687	24681:	contig of 1995 bp in length
*	24682	24781:	gap of unknown length
*	24782	26075:	contig of 1294 bp in length
*	26076	26175:	gap of unknown length
*	26176	28187:	contig of 2012 bp in length
*	28188	29287:	gap of unknown length
*	29288	29381:	contig of 1004 bp in length
*	29382	31317:	gap of unknown length
*	31318	31417:	contig of 1926 bp in length
*	31419	32960:	contig of 1543 bp in length
*	32961	33060:	gap of unknown length
*	33061	34513:	contig of 1453 bp in length
*	34514	34613:	gap of unknown length
*	34614	36216:	contig of 1603 bp in length
*	36217	36316:	gap of unknown length
*	36317	38470:	contig of 2154 bp in length
*	38471	38570:	gap of unknown length
*	38571	40918:	contig of 2348 bp in length
*	40919	41018:	gap of unknown length
*	41019	43402:	contig of 2384 bp in length
*	43403	43502:	gap of unknown length
*	43503	46402:	contig of 2900 bp in length
*	46403	46502:	gap of unknown length
*	46503	48711:	contig of 2209 bp in length
*	48712	48811:	gap of unknown length
*	48812	51671:	contig of 2760 bp in length
*	51672	51771:	gap of unknown length
*	51772	53561:	contig of 1910 bp in length
*	53562	53661:	gap of unknown length
*	53662	55602:	contig of 1921 bp in length
*	55603	55702:	gap of unknown length
*	55703	56938:	contig of 1236 bp in length
*	56939	57038:	gap of unknown length
*	57039	59030:	contig of 2052 bp in length
*	59041	59190:	gap of unknown length

FEATURES	source	location/Qualifiers	
BASE COUNT	20702 a 18794 c 18175 g 20853 t 5137 others		
ORIGIN	1..83661 /organism="Rattus norvegicus" /db_xref="taxon:10116" /clone="CH230-118h3"		
Query Match	83.2% Score 20.8; DB 2: Length 83661;		
Best Local Similarity	91.7% Pred. No. 1.8e+02;		
Matches	22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;		
QY	2 CATATATATCTGACATTCMAAACAA 25		
DB	30157 CATATATCTGACATTCMAAACAA 30180		
DEFINITION	AC105540 152557 bp DNA linear HTG 13-JUL-2002		
LOCUS	Rattus norvegicus clone CH230-232018, *** SEQUENCING IN PROGRESS		
AC105540	***, 70 unordered pieces.		
ACCESSION	AC105540		
VERSION	AC105540.2 GI:21736405		
KEYWORDS	HTG, HTGS, PHASE1.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
AUTHORS	1 (bases 1 to 152557) Muzny,D.M., Adams,C., Adic-Odoia,B., Ali-osman,F.R., Allen,C., Alldbrooks,S.T., Amaraltinge,H.C., Are,J.R., Ayale,M., Banks,T., Bardarika,J., Benton,J., Blumage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Daylla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hultj,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudash,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H., Locado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Madeshtwari,M., Magua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mashiney,E., Mcleod,M.P., Meador,M., Mei,G., Metzger,M., Mitter,G., Mitter,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neel,D., Nektsen,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenko,S., Oguh,M., Okunnu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,		

----- Project Information -----  
Center project name: H\_NH1296011  
----- Summary Statistics -----  
Sequencing vector: M13; 28%  
Sequencing vector: plasmid; 72%  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 206799 bases at least Q40  
Consensus quality: 207057 bases at least Q30  
Consensus quality: 207288 bases at least Q20  
Insert size: 217000; agarose-fp  
Insert size: 207414; sum-of-contigs  
Quality coverage: 10.57 in Q20 bases; agarose-fp  
Quality coverage: 11.06 in Q20 bases; sum-of-contigs  
----- NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1076: contig of 1076 bp in length  
\* 1077 1176: gap of unknown length  
\* 1177 2354: contig of 1178 bp in length  
\* 2355 2454: gap of unknown length  
\* 2455 207614: contig of 205160 bp in length.  
\* Location/Qualifiers  
1. 207614  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="UNK"  
/clone="RP11-1296011"  
1. 1076  
/note="assembly\_name:Contig21"  
1177. 2354  
/note="assembly\_name:Contig23"  
2455. 207614  
/note="assembly\_name:Contig27  
clone\_end:SP6  
vector\_side:right"  
BASE COUNT 62251 a 42536 c 43366 g 59199 t 202 others  
ORIGIN  
Query Match 85.6% Score 21.4; DB 2; Length 207614;  
Best Local Similarity 95.7%; Pred. No. 86;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Oy 3 ATATATCTCATTCATCAAAACAA 25  
|||||  
Db 153000 ATATATCTCATTCATCAAAACAA 153022  
-----  
RESULT 4  
AC123500 83661 bp DNA linear HTG 24-JUL-2002  
LOCUS AC123500  
DEFINITION Rattus norvegicus clone CH230-118H3, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*\*\* 41 unordered pieces.  
AC123500  
VERSION AC123500.2 GI:21909244  
KEYWORDS HTG; HTGS-PHASE1.  
SOURCE Rattus norvegicus.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 83661)  
Murray,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Albrooks,S.L., Amaralungu,H.C., Are,J.R., Ayale,M., Banks,T.,  
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,

Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dachorne,S.R., David,R.,  
Daya,M., Davis,C., Davy-Carroll,L., Deckerich,D.A.,  
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dihn,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
Homsli,F., Howard,S., Huber,J., Huliyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Joliver,S., Joudah,S.,  
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovicic,J., Kurstul,A., Landy,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lightarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,  
Lozado,R.J., Lu,X., Luchter,A., Luchter,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mayhew,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,  
Minerl,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Moser,M., Neal,D., Newton,J., Newton,S., Ogun,M., Okunolu,G.,  
Nguyen,N., Nickerson,E., Nwokenko,S., Ogun,M., Okunolu,G.,  
Oragunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peterson,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rivers,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Saverly,G.,  
Shaver,S., Scott,G., Shen,H., Shoochari,N., Sisson,I.,  
Sodergren,E., Sotaike,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telirod,D., Thomas,N., Thomas,S.,  
Usmani,K., Vazquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.  
Weinstock,G. and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 83661)  
Worley,K.C.  
Direct Submission  
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 83661)  
Worley,K.C.  
Direct Submission  
Submitted (24-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 19, 2002 this sequence version replaced gi:21240448.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GYRN  
Center clone name: CH230-118H3  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 43471 bases at least Q40  
Consensus quality: 46338 bases at least Q30  
Consensus quality: 48795 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 41 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is

```

repeat_region      1. .71
/rpt_family="ERV1"
repeat_region      77. .1621
/rpt_family="L1"
repeat_region      1628. .1825
/rpt_family="L1"
repeat_region      2549. .2861
/rpt_family="Alu"
repeat_region      2835. .2871
/rpt_family="A-rich"
repeat_region      3021. .3191
/rpt_family="MER1-type"
repeat_region      4341. .4712
/rpt_family="MALR"
repeat_region      5872. .5907
/rpt_family="AT-rich"
repeat_region      5893. .6190
/rpt_family="Alu"
repeat_region      6458. .6754
/rpt_family="Alu"
repeat_region      7331. .7597
/rpt_family="L1"
repeat_region      7575. .7623
/rpt_family="L1"
repeat_region      7709. .8020
/rpt_family="AT-rich"
repeat_region      8162. .8500
/rpt_family="L1"
repeat_region      8671. .8915
/rpt_family="L1"
repeat_region      8931. .10971
/rpt_family="L1"
repeat_region      10998. .11270
/rpt_family="Alu"
repeat_region      11244. .11331
/rpt_family="(GAA)n"
repeat_region      11322. .12313
/rpt_family="L1"
repeat_region      12304. .12443
/rpt_family="(TA)n"
repeat_region      12408. .12474
/rpt_family="ERV1"
repeat_region      12479. .12824
/rpt_family="L1"
repeat_region      12825. .13431
/rpt_family="L2"
repeat_region      14960. .15290
/rpt_family="MALR"
repeat_region      15853. .15918
/rpt_family="Alu"
repeat_region      15919. .17117
/rpt_family="ERV1"
repeat_region      17123. .17365
/rpt_family="Alu"
repeat_region      17580. .17584
/rpt_family="Alu"
misc_feature      a126f06.sl"
17963. .18256
/Note="similar to Homo sapiens EST AA781453 (NID:g2840784)
repeat_region      18416. .18500
/rpt_family="MALR"
repeat_region      18501. .18810
/rpt_family="Alu"
repeat_region      18785. .18830
/rpt_family="(TAA)n"
repeat_region      18811. .19118
/rpt_family="MALR"
repeat_region      19397. .19634
/rpt_family="AT-rich"
repeat_region      20939. .20967
/rpt_family="AT-rich"
repeat_region      22281. .22302
/rpt_family="AT-rich"

```

```

repeat_region      22588. .22885
/rpt_family="Alu"
repeat_region      23047. .23095
/rpt_family="AT-rich"
repeat_region      23425. .23530
/rpt_family="L2"
repeat_region      23972. .24318
/rpt_family="MER2-type"
repeat_region      25329. .26141
/rpt_family="L1"
repeat_region      26142. .26436
/rpt_family="Alu"
misc_feature      26531. .26387
/Note="similar to Homo sapiens EST R36044 (NID:g792945)
misc_feature      26336. .26368
/Note="similar to Homo sapiens EST AW881127
repeat_region      26437. .27763
/rpt_family="L1"
repeat_region      28812. .29038
/rpt_family="L1"
repeat_region      29038. .29062
/rpt_family="(TTA)n"
repeat_region      29082. .29313
/rpt_family="Alu"
misc_feature      29114. .29147
/Note="similar to Homo sapiens EST AW881127
repeat_region      29324. .29801
/rpt_family="L1"
misc_feature      29890. .30234
/Note="similar to Homo sapiens EST AW881127
(NID:g8043137)"
Query Match      85.6%; Score 21.4; DB 9; Length 175895;
Best Local Similarity 95.7%; Pred. No. 89;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      3 ATATATCTGACATTCGAAACAA 25
Db      28650 ATATATCTGACATTCGAAACAA 28672
RESULT 3
AC099800      207614 bp      DNA      linear      HTG 03-JAN-2002
LOCUS      Homo sapiens chromosome UNK clone RPL-1296C11, WORKING DRAFT
DEFINITION      AC099800
SEQUENCE      3 unordered pieces.
ACCESSION      AC099800
VERSION      AC099800.2 GI:1804257
KEYWORDS      HTG; HTGS-PHASE1; HTGS-DRAFT; HTGS-FULLTOP.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 207614)
AUTHORS      Waterston,R.H.
TITLE      The sequence of Homo sapiens clone
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 207614)
AUTHORS      Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (21-NOV-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT      On Jan 3, 2002 this sequence version replaced gi:17027321.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web Site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu

```



PDPAQGVITITGVDVDRARMLRRNCFTVEMALSKMPLHLEQAVHVRNPIONV  
 EMSEVDAPEKNGIATVITLRTAREKNOHLEIAKAKNELFPESAPENKFLHFTL  
 STYVVDVGLSSSTALMPYIERETITITPCYNNRNRGRIYIKVYGNIDVTLK  
 ARRYIMDLPISMCFNKNDMAPEPSVRNHNHIDEGITLTKMTPSYEPADLS  
 GEVPLNCASLRSKENFKITLYOKYLSKRFPIAPQPDIDNSITWHHSIPNFKN  
 FNNPCRELDGSGRRHRSSIASSSKRSYKSKOPSESSGGGSRHTRVRSSE  
 NSSVPIPMOPFTPHFAPMLTPHHMLKYVYLQOHOAOFLKGAAGLHPCGHIMEPP  
 PIYVDSFGARYSNITPFYSYIVQNNFNNGNCHL"

16033. 20426  
 /gene="T07F8.2"  
 /note="for a graphical representation of this gene see:  
 http://www.wormbase.org/db/seq/sequence?name=T07F8.2;class=Sequence"  
 join(16033. 16081,16586. 16697,16746. 16879,16933. 17322,  
 17739. 17866,18090. 18163,18212. 18319,20099. 20139,  
 20265. 20325,20372. 20426)  
 /gene="T07F8.2"  
 /note="exons 2 and 4 similar to transmembrane and  
 cytoplasmic domains of somatostatin receptor type 1"  
 /codon\_start=1  
 /product="Hypothetical protein T07F8.2"  
 /protein\_id="AAB38097.1"  
 /db\_xref="GI:1708759"  
 /translation="MPEIREIRPQISRYEDMSPYFAGFLYIAVLTGLPQINWY  
 CSVARSRKPKSLMSRSPSDRLRAYISLAVIDITVALLVPAHLFHPHMLDMS  
 CRAMEVLENSVKITSLTLCISLERITIRKPESEVRQFNAPDIGASIVGLAV  
 GAIIVQINSVSSDGLNCVRSRGKAIIPRASYLEAVLVLTLITLSISQIVHVV  
 RRKTRRARQANSRVRESIVNPEPYRMTAIVAGVHVVQPMPLMOEIPDN  
 TIOELTAGIRLRFSEFRDYSTRILFIATWLSMAGMIVIAVVRNRLRLISAV  
 AKTYFOSSATGEEKISSRAGSNINPQLRTITFCFTFNSQSTRKL"  
 21220. 23732  
 /gene="T07F8.1"  
 /note="for a graphical representation of this gene see:  
 http://www.wormbase.org/db/seq/sequence?name=T07F8.1;class=Sequence"  
 join(21220. 23253,23298. 23732)  
 /gene="T07F8.1"  
 /note="weak similarity in 5' end of exon 1 to C. elegans  
 myosin heavy chain B; 3' end of exon 1 similar to ZK686.4"  
 /codon\_start=1  
 /product="Hypothetical protein T07F8.1"

Query Match 85.6%; Score 21.4; DB 3; Length 23862;  
 Best Local Similarity 95.7%; Pred. NO. 1.3e+02;  
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CATATATATCTACATTCATAAACA 24  
 |||||  
 DB 14631 CATAAATCTACATTCATAAACA 14653

RESULT 2  
 AC064862 175895 bp DNA linear PRI 09-JAN-2002  
 LOCUS  
 DEFINITION Homo sapiens BAC clone RP11-368M16 from 7, complete sequence.  
 AC064862  
 ACCESSION  
 VERSION AC064862.6 GI:15145584  
 KEYWORDS  
 HTG.  
 SOURCE  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 175895)  
 Sulston, J.E. and Waterston, R.  
 Toward a complete human genome sequence  
 JOURNAL  
 MEDLINE  
 99063792  
 9847074  
 REFERENCE  
 2 (bases 1 to 175895)  
 Kruchowski, S., Kozlowski, A., Hawkins, M., Elliot, G. and Doeber, A.  
 The sequence of Homo sapiens BAC clone RP11-368M16  
 JOURNAL  
 Unpublished (2001)  
 3 (bases 1 to 175895)

AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 4 (bases 1 to 175895)  
 WATERSTON, R.H.  
 DIRECT SUBMISSION  
 Submitted (23-APR-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 5 (bases 1 to 175895)  
 WATERSTON, R.H.  
 DIRECT SUBMISSION  
 Submitted (09-AUG-2001) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 6 (bases 1 to 175895)  
 WATERSTON, R.H.  
 DIRECT SUBMISSION  
 Submitted (09-JAN-2002) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Aug 9, 2001 this sequence version replaced gi:14589722.  
 ----- Genome Center.  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc  
 Contact: sapiens@wustl.edu  
 ----- Summary Statistics  
 Center project name: H\_NH0368M16

NOTICE: This sequence may not represent the entire insert of this  
 clone. It may be shorter because we only sequence overlapping  
 clone sections once, or longer because we provide a small overlap  
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate  
 chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by sequence  
 from more than one subclone; and the assembly was confirmed by  
 restriction digest.

MAPPING INFORMATION:  
 The sequence of this clone was established as part of a mapping and  
 sequencing collaboration between the NHGRI Chromosome 17 Mapping  
 Project (Eric D. Green, Director), John D. McPherson in the  
 Department of Genetics (Washington University), and the Washington  
 University Genome Sequencing Center. For additional information  
 about the map position of this sequence, see  
 http://www.nhgri.nih.gov/DIR/GRB/CHR17, send  
 mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:  
 The RPCT-11 human BAC library was made from the blood of one male  
 donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,  
 Taten, M., Catanese, J.J., and de Jong, P.J. (1998) An improved  
 approach for construction of bacterial artificial chromosome  
 libraries. Genomics 51:1-8. The clone may be obtained either from  
 Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong  
 and coworkers at the Roswell Park Cancer Institute  
 (http://dpcpac.med.buffalo.edu)  
 VECTOR: pBACe3.6  
 NEIGHBORING SEQUENCE INFORMATION:  
 Actual start of this clone is at base position 1 of RP11-368M16;  
 actual end is at base position 175895 of RP11-368M16.

RP11-368M16 contains single plasmid regions from 51174 to 51278 and  
 80620 to 80630.

FEATURES  
 source  
 1. 175895  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="7"  
 /clone="RP11-368M16"  
 /clone\_11b="RPCT-11"

JOURNAL Science 282 (5396), 2012-2018 (1998)  
 MEDLINE 99069613  
 PUBMED 9851916  
 REFERENCE 2 (bases 1 to 23862)  
 AUTHORS Pauley, A.  
 TITLE The sequence of *C. elegans* cosmid T07F8  
 JOURNAL Unpublished (2001)  
 REFERENCE 3 (bases 1 to 23862)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-DEC-1996)  
 REFERENCE 4 (bases 1 to 23862)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-JUL-2001) Department of Genetics, Washington University, Genome Sequencing Center, 444 Forest Park Avenue, St. Louis, MO 63110, USA  
 REFERENCE 5 (bases 1 to 23862)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-MAY-2002) Department of Genetics, Washington University, Genome Sequencing Center, 444 Forest Park Avenue, St. Louis, MO 63110, USA

## COMMENT

Submitted by:  
 Genome Sequencing Center  
 Department of Genetics, Washington University  
 St. Louis, MO 63110, USA, and  
 Sanger Centre, Hinxton Hall  
 Cambridge CB10 1PO, England  
 email: r.w@nemastode.wustl.edu and jesus@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its analysis see:

<http://www.wormbase.org/db/seq/sequence?name=T07F8;class=Sequence>

## NEIGHBORING COSMID INFORMATION

The 5' cosmid is F10E7, 1000 bp overlap; the 3' cosmid is C27H5, 500 bp overlap. Actual start of this cosmid is at base position 1 of T07F8; actual end is at 12559 of C27H5.

## NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program GeneFinder (P. Green and L. Hillier, personal communication), the large scale BSR projects of Yujin Kohara ([http://www.ddbj.nig.ac.jp/c-elegans/html/CE\\_INDEX.html](http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html)) and The C. elegans ORFome cloning project (<http://wofid.dfcil.harvard.edu/>), similarity to other proteins from Blast analysis (<http://blast.wustl.edu/>), sequence conservation with *C. briggsae* using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual *C. elegans* Genbank submissions, and personal communications with *C. elegans* researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

## FEATURES

Source

1. 23862  
 /organism="Caenorhabditis elegans"  
 /strain="Bristol N2"  
 /db\_xref="taxon:6239"

/chromosome="II"  
 /clone="T07F8"  
 complement(2568..3987)  
 /gene="T07F8.4"  
 /note="For a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=T07F8.4;class=Sequence>"

## CDS

complement(join(2668..2845,2887..2948,3001..3353,3401..3596,3645..3755,3798..3854,3906..3987))

## CDS

/note="Contains similarity to Pfam domain: PF00249 (myb\_DNA-binding), Score=11.3, E-value=0.026, N=1; coded for by the following *C. elegans* cDNAs: YK132d3.5, YK132d3.3, YK479h9.3, YK479h9.5"

## CDS

/codon\_start=1  
 /product="Hypothetical protein T07F8.4"  
 /protein\_id="AAK38099.1"

## CDS

/db\_xref="GI:1708761"  
 /translation="MDEGSRSSITDEEDVDNDNGEDYEATIEEELLEDGDEDE  
 LKLEDEGMSIELOKXGYRPAENSSANNEDTEAEENVGSSAAEPHDSTIFD  
 MGSMSGFDDEDDDYAPPPPMKRSIRVDPLQADVPLENEATVESAREDDTIAT  
 IDOTNPSDEVINDYLNKDYGLKRAHDQPPAGTSGRDEEDALCALYSPNETEAK  
 ESPPPHINAPFRTVRSDDALGDESEKAPESSELELXGDFSLIRRLPLPYKVGELI  
 EYVQKLTGPGYRVNRDAPQAPVYQPHLSAHHQVQALDQNGCTGFVASESP  
 STSEPTLTN"

## CDS

4614..9003  
 /gene="T07F8.3"  
 /note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=T07F8.3a;class=Sequence>"

## CDS

join(4614..4738,4787..4868,4910..5020,5066..5136,5179..5344,5389..5515,5560..5681,5728..5805,5850..5987,6035..6227,6326..6424,6473..6641,6784..6906,6951..7019,7355..7432,7477..7716,7764..8190,8310..8591,8820..9003)

## CDS

/note="coded for by the following *C. elegans* cDNAs: YK609d3.3, cm13g9, YK29c11.5, YK163e6.5, YK35f11.5, YK609d3.3"

## CDS

/codon\_start=1  
 /product="Hypothetical protein T07F8.3a"  
 /protein\_id="AAK7629.1"

## CDS

/translation="KSGESRKHNDHSHNPPVRSAYVEDADTPTQLQMAENASAKL  
 FVSSALKDITVNEPFTYHROQSAQMAEDNORQVSYNTKRAHILHQLAQGLPIS  
 NIPMIEIPRTVLNMFEEQYSLMTSDGDENAVSIAETNTLILPDRSGVGT  
 PDPEAOVITLVYFGCDVDRALMLRNRCHEVTVMALSKKMPJHELOAVRQNDLVN  
 EMGFDAPEKNGIIVTTLITAREKQHELEAKRLNELIFESPAPNNFTLHFL  
 STYVDOVLSSSTQALMPYIERETTLIISPCYNNRNRGVIYKIVGNNDVILK  
 ARYIMDLPIISMCPNKNMDMAPEPSRPNHMTIDSGILTKTPESVYEPADLS  
 GEPVINCASRSKEPNIKKLYTAYOKVLSKRPFLAPQNDYNSIMHNSLPANFLN  
 FNNPKCELEDSNGRRHRSSIASRSKHSYISKQKQFSESGGFSRSHTRVSESE  
 NSSTVPTMQPPTPHFAPMILTPHHMLKYYVLOHQAOQFLGAGALHGHINFP  
 PLIVDSFVSALFPADPVYFDGPFYVHGJLFPVNEADHNRHRSPLSTQEIHKPS  
 RMNGNRSSSTGYSYPTPRQRYVEQVREDLRSHIGSRRTSVNCDQNVSMHQG  
 YEROYPROHRLQDKDOQKMTSGRQDINHSRTINHYRDYRNENYDFVAGSGPKR  
 SPLEDOYLOLMTHLKLSNDVLDHEKTLHMSPPMNSDVTYASGQNDMDGQF  
 QRLSNANINESRRPRYTSCTFEKQGSRRYIDSCGASVYDHASTHOSRSDYDGRK  
 VGNNGYKTLLEPRAREKQYKISLEHTKYSNERGDEKSEKSNETSLSGSRQYID  
 PMKRNYYILNRRDSNSFFYLHQFSLYNNPDIAYIRIHOFLSYSPSLCISRLISLP  
 RDIYD"

## CDS

join(4614..4738,4787..4868,4910..5020,5066..5136,5179..5344,5389..5515,5560..5681,5728..5805,5850..5987,6035..6227,6326..6424,6473..6641,6784..6906,6951..7108)

## CDS

/gene="T07F8.3"  
 /note="coded for by the following *C. elegans* cDNAs: YK345b5.3, YK719f7.3, YK345b5.5, YK19f7.5"

## CDS

/codon\_start=1  
 /product="Hypothetical protein T07F8.3b"  
 /protein\_id="AAK77630.1"

## CDS

/db\_xref="GI:15011789"  
 /translation="KGEQSEKDHDAHSYNPVRSAYVEDADTPTQLQMAENASAKL  
 FVSSALKDITVNEPFTYHROQSAQMAEDNORQVSYNTKRAHILHQLAQGLPIS  
 NIPMIEIPRTVLNMFEEQYSLMTSDGDENAVSIAETNTLILPDRSGVGT

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:07:09 ; Search time 88.7891 seconds

(without alignments)  
8194.363 Million cell updates/sec

Title: US-09-477-082-32

Perfect score: 25  
Sequence: 1 ccatatcatctacattcaaaacaa 25

Scoring table: IDENTITY-NUC  
Gapop 10.0 ; Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_da: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vi: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htg\_mus: \*  
34: em\_htg\_pln: \*  
35: em\_htg\_rtd: \*  
36: em\_htg\_mam: \*  
37: em\_htg\_vrt: \*  
38: em\_sy: \*  
39: em\_htgo\_hum: \*  
40: em\_htgo\_mus: \*  
41: em\_htgo\_other: \*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	21.4	85.6	23862	3	U80954	U80954 Caenorhabditis
2	21.4	85.6	175895	2	AC064862	AC064862 Homo sapi
3	21.4	85.6	207614	2	AC099800	AC099800 Homo sapi
4	20.8	83.2	83661	2	AC123500	AC123500 Rattus no
5	20.8	83.2	152557	2	AC105540	AC105540 Rattus no
6	20.8	83.2	188458	2	AC092273	AC092273 Homo sapi
7	20.4	81.6	71000	8	SPB887	AL032654 S. pombe c
8	20.4	81.6	127682	9	AL137861	AL137861 Human DNA
9	20.4	81.6	177283	8	AP003376	AP003376 Oryza sat
10	20.4	81.6	226726	2	AC115717	AC115717 Mus muscu
11	20.4	81.6	270987	2	AC122828	AC122828 Mus muscu
12	20.2	80.8	17280	6	AX251530	AX251530 Sequence
13	20.2	80.8	25539	9	AL161775	AL161775 Human DNA
14	20.2	80.8	114016	9	AL365396	AL365396 Homo sapi
15	20.2	80.8	140137	2	AC027611	AC027611 Homo sapi
16	20.2	80.8	183786	2	AL807741	AL807741 Mus muscu
17	20.2	80.8	185237	2	AC111638	AC111638 Rattus no
18	20.2	80.8	252079	2	AC115349	AC115349 Rattus no
19	20.2	80.0	68726	8	AB017063	AB017063 Arabidops
20	20.0	80.0	222323	2	AC091254	AC091254 Mus muscu
21	19.8	79.2	5474	6	AX458550	AX458550 Sequence
22	19.8	79.2	46517	5	AL672060	AL672060 Zebrafish
23	19.8	79.2	103194	8	AC006223	AC006223 Arabidops
24	19.8	79.2	119958	2	AC104706	AC104706 Oryza sat
25	19.8	79.2	122680	2	AC130231	AC130231 Rattus no
26	19.8	79.2	123182	9	AC099339	AC099339 Homo sapi
27	19.8	79.2	129420	8	AC087547	AC087547 Oryza sat
28	19.8	79.2	141041	8	AC087550	AC087550 Oryza sat
29	19.8	79.2	156533	2	AC117070	AC117070 Dictyoste
30	19.8	79.2	156721	9	AL139389	AL139389 Human DNA
31	19.8	79.2	162410	2	AL133265	AL133265 Homo sapi
32	19.8	79.2	166978	10	AC127431	AC127431 Mus muscu
33	19.8	79.2	168182	2	AC023799	AC023799 Homo sapi
34	19.8	79.2	178815	2	AC100883	AC100883 Mus muscu
35	19.8	79.2	186044	9	AC009803	AC009803 Homo sapi
36	19.8	79.2	194005	2	AC130723	AC130723 Mus muscu
37	19.8	79.2	201500	2	AC103946	AC103946 Mus muscu
38	19.8	79.2	203335	2	AC122056	AC122056 Mus muscu
39	19.8	79.2	203713	2	AC084725	AC084725 Mus muscu
40	19.8	79.2	209320	2	AC016884	AC016884 Homo sapi
41	19.8	79.2	223940	2	AC087567	AC087567 Mus muscu
42	19.4	77.6	65869	2	AC113552	AC113552 Homo sapi
43	19.4	77.6	106866	2	AC126007	AC126007 Medicago
44	19.4	77.6	133181	9	AC100854	AC100854 Homo sapi
45	19.4	77.6	166114	9	AC083837	AC083837 Homo sapi

#### ALIGNMENTS

RESULT 1  
U80954  
LOCUS U80954 23862 bp DNA linear INV 23-MAY-2002  
DEFINITION Caenorhabditis elegans cosmid i07f8, complete sequence.  
ACCESSION U80954  
VERSION U80954.1 GI:1708757  
KEYWORDS HTG.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
Eukaryote; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditidae; Rhabditidae; Felodermatidae; Caenorhabditis.  
REFERENCE 1 (bases 1 to 23862)  
AUTHORS Waterston, R.  
TITLE Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium

Pred. No. is the number of results predicted by chance to have a

Best Local Similarity 89.5%; Pred No. 4.2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGGGATTGGAGATTGTG 20  
|||  
Db 715 AGCGGATTGTGATGTG 733

RESULT 15  
US-09-873-880-29/C  
; Sequence 29, Application US/09873880  
; Patent No. US20020123118A1  
; GENERAL INFORMATION:  
; APPLICANT: Seralit, Vincent  
; APPLICANT: Falco, S. Carl  
; APPLICANT: Allen, Stephen M.  
; TITLE OF INVENTION: GLYCINE METABOLISM ENZYMES  
; FILE REFERENCE: B81192 US CIP  
; CURRENT APPLICATION NUMBER: US/09/873,880  
; CURRENT FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: 09/363,321  
; PRIOR FILING DATE: July 28, 1999  
; PRIOR APPLICATION NUMBER: 60/094,839  
; PRIOR FILING DATE: July 31, 1998  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 29  
; LENGTH: 1362  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-873-880-29

Query Match 75.2%; Score 15.8; DB 10; Length 1362;  
Best Local Similarity 89.5%; Pred. No. 4.2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AGGGGATTGGAGATTGTG 20  
|||  
Db 360 AGGGGATTGGAGATTGCG 342

Search completed: July 8, 2003, 06:04:29  
Job time : 25.6338 secs

Best Local Similarity 85.7%; Pred. No. 3.2e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATTGTGA 21  
Db 2829 TAGGGATTGGTGTGATTGAGA 2849

## RESULT 11

US-10-172-086-11  
Sequence 11, Application US/10172086  
Publication No. US20030113750A1  
GENERAL INFORMATION:  
APPLICANT: Epigenomics AG  
TITLE OF INVENTION: Method and nucleic acids for the differentiation  
TITLE REFERENCE: of prostate tumors  
CURRENT APPLICATION NUMBER: US/10/172,086  
CURRENT FILING DATE: 2002-06-13  
NUMBER OF SEQ ID NOS: 116  
SEQ ID NO 11  
LENGTH: 5310  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-172-086-11

## Query Match

Best Local Similarity 85.7%; Pred. No. 3.2e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATTGTGA 21  
Db 3989 TAGGGGTGTGGAGATTATGA 4009

## RESULT 12

US-09-873-880-7/c  
Sequence 7, Application US/09873880  
Patent No. US20020123118A1  
GENERAL INFORMATION:  
APPLICANT: Sewalt, Vincent  
APPLICANT: Faico, S. Carl  
APPLICANT: Allen, Stephen M.  
TITLE OF INVENTION: GLYCINE METABOLISM ENZYMES  
FILE REFERENCE: B81192 US CIP  
CURRENT APPLICATION NUMBER: US/09/873,880  
CURRENT FILING DATE: 2001-06-04  
PRIOR APPLICATION NUMBER: 09/363,321  
PRIOR FILING DATE: July 28, 1999  
PRIOR APPLICATION NUMBER: 60/094,839  
PRIOR FILING DATE: July 31, 1998  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 7  
LENGTH: 495  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (382)  
OTHER INFORMATION: n=A, C, G, or T  
NAME/KEY: unsure  
LOCATION: (434)  
OTHER INFORMATION: n=A, C, G, or T  
US-09-873-880-7

## Query Match

Best Local Similarity 89.5%; Pred. No. 3.8e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGGATTGGAGATTGTG 20

Db 353 AGGGATTGGAGATTGCG 335

## RESULT 13

US-10-001-879-38/c  
Sequence 38, Application US/10001879  
Patent No. US20020127237A1  
GENERAL INFORMATION:  
APPLICANT: Salceda, Susana  
APPLICANT: Macina, Roberto  
APPLICANT: Recipon, Hervé  
APPLICANT: Cafferey, Robert  
APPLICANT: Ali, Shujath  
APPLICANT: Sun, Yongming  
APPLICANT: Liu, Chenghua  
TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes a  
FILE REFERENCE: DEX-0281  
CURRENT APPLICATION NUMBER: US/10/001,879  
CURRENT FILING DATE: 2001-11-20  
PRIOR APPLICATION NUMBER: 60/252,188  
PRIOR FILING DATE: 2000-11-21  
NUMBER OF SEQ ID NOS: 201  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 38  
LENGTH: 920  
TYPE: DNA  
ORGANISM: Homo sapien  
US-10-001-879-38

## Query Match

Best Local Similarity 89.5%; Pred. No. 4.1e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGGGATTGGAGATTGTGA 21  
Db 692 GGGGATGAGAGATTGTGA 674

## RESULT 14

US-09-738-626-2066  
Sequence 2066, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIICHI  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 2066  
LENGTH: 1149  
TYPE: DNA  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-2066

## Query Match

Best Local Similarity 75.2%; Score 15.8; DB 9; Length 1149;

```

; APPLICANT: Foy, Theresa M.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 210121, 479C1
; CURRENT APPLICATION NUMBER: US/09/757,417
; CURRENT FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 422
; TYPE: DNA
; ORGANISM: Oryctolagus cuniculus
US-09-757-417-23

Query Match      77.1%; Score 16.2; DB 10; Length 422;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TAGGGGATTGGAGATTGTGA 21
      1 ||||| ||||| ||||| |||||

DB      280 TCGGGGTTTGGAGATGTGA 260

RESULT 7
US-09-918-995-33366
; Sequence 33366, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33366
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-33366

Query Match      77.1%; Score 16.2; DB 9; Length 430;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TAGGGGATTGGAGATTGTGA 21
      1 ||||| ||||| ||||| |||||

DB      347 TAGGAGATTGGAAATTCTGA 367

RESULT 8
US-09-918-995-34536
; Sequence 34536, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34536
; LENGTH: 475
; TYPE: DNA

```

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(475)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-34536

Query Match      77.1%; Score 16.2; DB 9; Length 475;
Best Local Similarity 85.7%; Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TAGGGGATTGGAGATTGTGA 21
      1 ||||| ||||| ||||| |||||

DB      442 TAGGGGAAATGGAGACTGTGA 462

RESULT 9
US-09-938-842A-4643/C
; Sequence 4643, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kleps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SRRIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4643
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4643

Query Match      77.1%; Score 16.2; DB 9; Length 2000;
Best Local Similarity 85.7%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TAGGGGATTGGAGATTGTGA 21
      1 ||||| ||||| ||||| |||||

DB      223 TAGGAGATTGGAGTTGTGA 203

RESULT 10
US-10-172-086-42
; Sequence 42, Application US/10172086
; Publication No. US20030113750A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Method and nucleic acids for the differentiation
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/172,086
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 42
; LENGTH: 5118
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-172-086-42

Query Match      77.1%; Score 16.2; DB 9; Length 5118;

```

```
; CURRENT FILING DATE: 2001-10-09
; NUMBER OF SEQ ID NOS: 43
; SEQ ID NO 11
; LENGTH: 21000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-975-123-11

Query Match          80.0%; Score 16.8; DB 9; Length 21000;
Best Local Similarity 90.0%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 AGGGGATTGGAGATTGTA 21
        |||||
Db       1787 AGGGGATTGGAGATTGTA 1768

RESULT 3
US-09-815-242-7388
; Sequence 7388, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELTRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7388
; LENGTH: 981
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(981)
US-09-815-242-7388

Query Match          78.1%; Score 16.4; DB 10; Length 981;
Best Local Similarity 94.4%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AGGGGATTGGAGATTGT 19
        |||||
Db       810 AGGGGATTGGAGATTGT 827

RESULT 4
US-10-102-524-97
; Sequence 97, Application US/10102524
```

```
; Publication No. US20030109434A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Gaiger, Alexander
; APPLICANT: Gordon, Brian
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 210121.572
; CURRENT APPLICATION NUMBER: US/10/102,524
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 1863
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 97
; LENGTH: 186
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 15, 51
; OTHER INFORMATION: n = A,T,C or G
US-10-102-524-97

Query Match          77.1%; Score 16.2; DB 9; Length 186;
Best Local Similarity 85.7%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TAGGGATTGGAGATTGTA 21
        |||||
Db       114 TTGGGATTGGGAGATTGTA 134

RESULT 5
US-10-042-945-23/C
; Sequence 23, Application US/10042945
; Publication No. US20030045468A1
; GENERAL INFORMATION:
; APPLICANT: Flinn, Steven P.
; APPLICANT: Foy, Teresa M.
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Wang, Aljun
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Sutherland, R. Alec
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY,
; FILE REFERENCE: 210121.479C3
; CURRENT APPLICATION NUMBER: US/10/042,945
; CURRENT FILING DATE: 2002-01-08
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 422
; TYPE: DNA
; ORGANISM: Oryctolagus cuniculus
US-10-042-945-23

Query Match          77.1%; Score 16.2; DB 9; Length 422;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TAGGGATTGGAGATTGTA 21
        |||||
Db       280 TCGGGATTGGAGATTGTA 260

RESULT 6
US-09-757-417-23/C
; Sequence 23, Application US/09757417
; Patent No. US20020082216A1
; GENERAL INFORMATION:
; APPLICANT: Fanger, Gary R.
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:57:09 ; Search time 10.6338 Seconds  
(without alignments)  
3082.690 Million cell updates/sec

Title: US-09-477-082-31

Perfect score: 21  
Sequence: 1 tagggagattgagattgtga 21

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications.NA.\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCIT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCITUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.8	84.8	436	10	US-09-878-574-4901
2	16.8	80.0	21000	9	US-09-975-123-11
3	16.4	78.1	981	10	US-09-815-242-7388
4	16.2	77.1	186	9	US-10-102-824-97
5	16.2	77.1	422	9	US-10-042-945-23
6	16.2	77.1	422	10	US-09-757-417-23
7	16.2	77.1	470	9	US-09-918-995-33366
8	16.2	77.1	475	9	US-09-918-995-33366
9	16.2	77.1	2000	9	US-09-838-842A-4643
10	16.2	77.1	5118	9	US-10-172-086-42
11	16.2	77.1	5310	9	US-10-172-086-41
12	15.8	75.2	495	10	US-09-873-880-7
13	15.8	75.2	920	12	US-10-001-879-38
14	15.8	75.2	1149	9	US-09-738-626-2066
15	15.8	75.2	1362	10	US-09-873-880-29
16	15.8	75.2	81826	9	US-10-175-523-197
17	15.8	75.2	3509400	9	US-09-738-626-1
18	15.4	73.3	877	9	US-09-770-445-539
19	15.4	73.3	7770	10	US-09-991-630-23

C 20	15.4	73.3	684973	10	US-09-263-959-1	Sequence 1, Appl
C 21	15.2	72.4	211	10	US-09-867-701-7906	Sequence 7906, Ap
C 22	15.2	72.4	229	10	US-09-867-701-7696	Sequence 7696, Ap
C 23	15.2	72.4	257	9	US-10-040-739-122	Sequence 122, App
C 24	15.2	72.4	336	9	US-09-796-692-2706	Sequence 2706, Ap
C 25	15.2	72.4	336	9	US-10-040-862-2706	Sequence 2706, Ap
C 26	15.2	72.4	348	9	US-09-736-457-1382	Sequence 1382, Ap
C 27	15.2	72.4	348	9	US-09-902-941-1382	Sequence 1382, Ap
C 28	15.2	72.4	348	9	US-09-849-626-1382	Sequence 1382, Ap
C 29	15.2	72.4	348	9	US-10-017-754-1382	Sequence 1382, Ap
C 30	15.2	72.4	372	10	US-09-878-574-1079	Sequence 1079, Ap
C 31	15.2	72.4	382	10	US-09-878-574-2535	Sequence 2535, Ap
C 32	15.2	72.4	425	10	US-09-860-352-2921	Sequence 2921, Ap
C 33	15.2	72.4	473	10	US-09-864-761-14436	Sequence 14436, A
C 34	15.2	72.4	473	10	US-09-783-590-5415	Sequence 5415, Ap
C 35	15.2	72.4	481	9	US-09-907-969-46	Sequence 46, Appl
C 36	15.2	72.4	481	10	US-09-884-441-46	Sequence 46, Appl
C 37	15.2	72.4	486	10	US-09-864-761-5386	Sequence 5386, Ap
C 38	15.2	72.4	510	9	US-09-918-995-2254	Sequence 2254, Ap
C 39	15.2	72.4	518	9	US-09-770-961-67	Sequence 67, Appl
C 40	15.2	72.4	528	8	US-08-945-038-9	Sequence 9, Appl
C 41	15.2	72.4	580	10	US-09-864-761-21927	Sequence 21927, A
C 42	15.2	72.4	871	9	US-09-989-442-45	Sequence 45, Appl
C 43	15.2	72.4	1056	9	US-09-938-842A-2129	Sequence 2129, Ap
C 44	15.2	72.4	1080	9	US-09-988-442-80	Sequence 80, Appl
C 45	15.2	72.4	1080	9	US-10-074-045-29	Sequence 29, Appl

#### ALIGNMENTS

RESULT 1  
US-09-878-574-4901/C  
; Sequence 4901, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Birm, Joseph R.  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Thompson, Michael D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878, 574  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/333, 535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 4901  
; LENGTH: 436  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: LIB3028-054-01-B1-B3  
US-09-878-574-4901

Query Match  
Best Local Similarity 84.8%; Score 17.8; DB 10; Length 436;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGATTGGAGATTGTGA 21  
1 |||||  
Db 170 TTGGGATTGGAGATTGTGA 150

RESULT 2  
US-09-975-123-11/C  
; Sequence 11, Application US/09975123  
; Publication No. US20030087857A1  
; GENERAL INFORMATION:  
; APPLICANT: Susan M. Freier  
; TITLE OF INVENTION: ANTISENSE MODULATION OF INSULIN-LIKE GROWTH FACTOR BINDING PRO  
; FILE REFERENCE: RTS-0253  
; CURRENT APPLICATION NUMBER: US/09/975,123

Fri Jul 11 09:24:03 2003

us-09-477-082-31.rni

Page 8

Job time : 18.0371 secs

---

QY 1 TAGGGATTGGAGATTGTG 20  
 |||| |||| || ||||  
 Db 1078 TAGGCATTGTAGATGTG 1059

## RESULT 14

US-09-173-891-5/C  
 ; Sequence 5, Application US/09173891  
 ; Patent No. 607937  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Payne, Jewel  
 ; APPLICANT: Kennedy, M. Keith  
 ; APPLICANT: Randall, John Brooks  
 ; APPLICANT: Meier, Henry  
 ; APPLICANT: Uick, Heidi Jane  
 ; APPLICANT: Foncerrada, Luis  
 ; APPLICANT: Schmeff, H. Ernest  
 ; APPLICANT: Schwab, George E.  
 ; APPLICANT: Fu, Jenny  
 ; TITLE OF INVENTION: No. 6077937e1 Bacillus thuringiensis Toxins Active  
 ; TITLE OF INVENTION: Against Hymenopteran Pests  
 ; NUMBER OF SEQUENCES: 51  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: David R. Saliwanchik  
 ; STREET: 2421 N.W. 41st Street, Suite A-1  
 ; CITY: Gainesville  
 ; STATE: FL  
 ; COUNTRY: USA  
 ; ZIP: 32606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/173,891  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/158,232  
 ; FILING DATE:  
 ; APPLICATION NUMBER: US 07/687,980  
 ; FILING DATE: 22-MAY-1992  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/797,645  
 ; FILING DATE: 25-NOV-1991  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/703,977  
 ; FILING DATE: 22-MAY-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Saliwanchik, David R.  
 ; REGISTRATION NUMBER: 31,794  
 ; REFERENCE/DOCKET NUMBER: M/SCU104.C1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 904-375-8100  
 ; TELEFAX: 904-372-5800  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3771 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Bacillus thuringiensis  
 ; INDIVIDUAL ISOLATE: 33f2  
 ; IMMEDIATE SOURCE:  
 ; CLONE: E. coli NM522(pMTC316) B-18785  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 4..24

OTHER INFORMATION: /function= "oligonucleotide  
 ; OTHER INFORMATION: hybridization probe"  
 ; OTHER INFORMATION: /product= "GCA/T ACA/T TTA AAT GAA GTA/T TAT"  
 ; OTHER INFORMATION: /standard\_name= "probe a"  
 ; OTHER INFORMATION: /note= "Probe A"  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 13..33  
 ; OTHER INFORMATION: /function= "oligonucleotide  
 ; OTHER INFORMATION: hybridization probe"  
 ; OTHER INFORMATION: /product= "AAT GAA GTA/T TAT CCA/T GTA/T AAT"  
 ; OTHER INFORMATION: /standard\_name= "Probe B"  
 ; OTHER INFORMATION: /label= probe-b  
 ; OTHER INFORMATION: /note= "probe b"

## US-09-173-891-5

Query Match 72.4% Score 15.2; DB 3; Length 3771;  
 Best Local Similarity 85.0% Pred. No. 2.2e+02;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGATTGGAGATTGTG 20  
 |||| |||| || ||||  
 Db 1078 TAGGCATTGTAGATGTG 1059

## RESULT 15

US-09-076-137-5/C  
 ; Sequence 5, Application US/09076137B  
 ; Patent No. 6166195  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schmeff, Harry E.  
 ; APPLICANT: Schwab, George E.  
 ; APPLICANT: Payne, Jewel M.  
 ; APPLICANT: Narva, Kenneth E.  
 ; APPLICANT: Foncerrada, Luis  
 ; TITLE OF INVENTION: No. 6166195e1 Nematode-Active Toxins and Genes Which Code  
 ; TITLE OF INVENTION: Therefor  
 ; FILE REFERENCE: MA-20CCCD2  
 ; CURRENT APPLICATION NUMBER: US/09/076,137B  
 ; EARLIER FILING DATE: 1998-05-12  
 ; PRIOR APPLICATION NUMBER: 08/316,301  
 ; EARLIER FILING DATE: 1994-09-30  
 ; NUMBER OF SEQ ID NOS: 42  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 5  
 ; LENGTH: 3771  
 ; TYPE: DNA  
 ; ORGANISM: Bacillus thuringiensis  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (4)..(24)  
 ; OTHER INFORMATION: /function= "oligonucleotide hybridization probe"  
 ; OTHER INFORMATION: /product= "GCA/T ACA/T TTA AAT GAA GTA/T TAT"  
 ; OTHER INFORMATION: /standard\_name= "probe a" /note= "Probe A"  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (13)..(33)  
 ; OTHER INFORMATION: /function= "oligonucleotide hybridization probe"  
 ; OTHER INFORMATION: /product= "AAT GAA GTA/T TAT CCA/T GTA/T AAT"  
 ; OTHER INFORMATION: /standard\_name= "Probe B" /label= probe-b /note=  
 ; OTHER INFORMATION: "probe b"  
 ; US-09-076-137-5

Query Match 72.4% Score 15.2; DB 4; Length 3771;  
 Best Local Similarity 85.0% Pred. No. 2.2e+02;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGATTGGAGATTGTG 20  
 |||| |||| || ||||  
 Db 1078 TAGGCATTGTAGATGTG 1059

Search completed: July 8, 2003, 05:56:57

FILED DATE: 12-AUG-1987  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/830,050  
FILING DATE: 31-JAN-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lloyd, Jeff  
REGISTRATION NUMBER: 35,589  
REFERENCE/DOCKET NUMBER: MA20CCDD1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 5 (PS33F2):  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3771 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
IMMEDIATE SOURCE: 33f2  
CLONE: E. coli NM522(pMYC 2316) B-18785  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 4..24  
OTHER INFORMATION: /function= oligonucleotide  
OTHER INFORMATION: /hybridization probe  
OTHER INFORMATION: /product= GCA/T ACA/T TTA AAT GAA GTA/T TAT  
OTHER INFORMATION: /standard\_name= probe a  
OTHER INFORMATION: /note= probe a  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 13..33  
OTHER INFORMATION: /function= oligonucleotide  
OTHER INFORMATION: /hybridization probe  
OTHER INFORMATION: /product= AAT GAA GTA/T TAT CCA/T GTA/T AAT  
OTHER INFORMATION: /standard\_name= Probe B  
OTHER INFORMATION: /label= probe-b  
OTHER INFORMATION: /note= probe b  
US-08-316-301a-5  
Query Match 72.4%; Score 15.2; DB 1; Length 3771;  
Best Local Similarity 85.0%; Pred. No. 2.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
OY 1 TAGGGGATTGGAGATTGTG 20  
DB 1078 TAGCGATTGTGTAAGATG 1059  
RESULT 13  
US-08-611-928-5/C  
Sequence 5, Application US/08611928  
Patent No. 5824/92  
GENERAL INFORMATION:  
APPLICANT: Payne, Jewel  
APPLICANT: Kennedy, M. Keith  
APPLICANT: Randall, John Brooks  
APPLICANT: Meier, Henry  
APPLICANT: Ulick, Heidi Jane  
APPLICANT: Forcetrada, Luis  
APPLICANT: Schnepf, H. Ernest  
APPLICANT: Schwab, George E.  
APPLICANT: Fu, Jenny  
TITLE OF INVENTION: No. 5824/92a1 Bacillus thuringiensis Toxins Active  
TITLE OF INVENTION: Against Hymenopteran Pests  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:

ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,928  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/158,232  
FILING DATE: 24-NOV-1993  
APPLICATION NUMBER: US 07/887,980  
FILING DATE: 22-MAY-1992  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/797,645  
FILING DATE: 25-NOV-1991  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/703,977  
FILING DATE: 22-MAY-1991  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/SCJ104.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3771 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
INDIVIDUAL ISOLATE: 33f2  
IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMYC2316) B-18785  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 4..24  
OTHER INFORMATION: /function= oligonucleotide  
OTHER INFORMATION: /hybridization probe  
OTHER INFORMATION: /product= GCA/T ACA/T TTA AAT GAA GTA/T TAT  
OTHER INFORMATION: /standard\_name= "probe a"  
OTHER INFORMATION: /note= "probe a"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 13..33  
OTHER INFORMATION: /function= oligonucleotide  
OTHER INFORMATION: /hybridization probe  
OTHER INFORMATION: /product= AAT GAA GTA/T TAT CCA/T GTA/T AAT  
OTHER INFORMATION: /standard\_name= "Probe B"  
OTHER INFORMATION: /label= probe-b  
OTHER INFORMATION: /note= "probe b"  
US-08-611-928-5  
Query Match 72.4%; Score 15.2; DB 1; Length 3771;  
Best Local Similarity 85.0%; Pred. No. 2.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

LOCATION: 13..33  
OTHER INFORMATION: /function="oligonucleotide"  
OTHER INFORMATION: hybridization probe"  
OTHER INFORMATION: /product="AAT GAA GTA/T TAT CCA/T GTA/T AAT"  
OTHER INFORMATION: /standard\_name="Probe B"  
OTHER INFORMATION: /label="probe-b"  
OTHER INFORMATION: /note="probe b"  
US-08-158-232-5

Query Match  
Best Local Similarity 85.0%; Pred No. 2.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

72.4%; Score 15.2; DB 1; Length 3771;  
1 TAGGGGATTGGAGATTGTG 20  
1078 TAGGGGATTGTGAGATTGTG 1059

RESULT 11  
US-08-304-626-5/c  
Sequence 5, Application US/08304626  
Patent No. 5616495  
GENERAL INFORMATION:  
APPLICANT: Payne, Jewel M.  
APPLICANT: Kennedy, M. Keith  
APPLICANT: Randall, John Brooks  
APPLICANT: Meier, Henry  
APPLICANT: Ulick, Heidi Jane  
APPLICANT: Foncerra, Luis  
APPLICANT: Schaeff, Harry E.  
APPLICANT: Schwab, George E.  
TITLE OF INVENTION: No. 5616495el Bacillus thuringiensis Isolates  
TITLE OF INVENTION: Active Against Hymenopteran Pests and Genes Encoding  
TITLE OF INVENTION: Hymenopteran Active Toxins  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Salivanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/304,626  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/887,980  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Salivanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/SCJ 104  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ. ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3771 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
INDIVIDUAL ISOLATE: 33F2

IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMKC2316) B-18785  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 4..24  
OTHER INFORMATION: /function="oligonucleotide"  
OTHER INFORMATION: hybridization probe"  
OTHER INFORMATION: /product="GCA/T ACA/T TTA AAT GAA GTA/T TAT"  
OTHER INFORMATION: /standard\_name="probe a"  
OTHER INFORMATION: /note="probe A"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 13..33  
OTHER INFORMATION: /function="oligonucleotide"  
OTHER INFORMATION: hybridization probe"  
OTHER INFORMATION: /product="AAT GAA GTA/T TAT CCA/T GTA/T AAT"  
OTHER INFORMATION: /standard\_name="Probe B"  
OTHER INFORMATION: /label="probe-b"  
OTHER INFORMATION: /note="probe b"  
US-08-304-626-5

Query Match  
Best Local Similarity 85.0%; Pred No. 2.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

72.4%; Score 15.2; DB 1; Length 3771;  
1 TAGGGGATTGGAGATTGTG 20  
1078 TAGGGGATTGTGAGATTGTG 1059

RESULT 12  
US-08-316-301A-5/c  
Sequence 5, Application US/08316301A  
Patent No. 5753492  
GENERAL INFORMATION:  
APPLICANT: Schaeff, Harry E.  
APPLICANT: Schwab, George E.  
APPLICANT: Payne, Jewel M.  
APPLICANT: Narva, Kenneth E.  
APPLICANT: Foncerra, Luis  
TITLE OF INVENTION: No. 5753492el Nematode-Active Toxins and Genes  
TITLE OF INVENTION: Which Code Therefor  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Salivanchik & Salivanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/316,301A  
FILING DATE: 30-SEP-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/871,510  
FILING DATE: 23-APR-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/693,018  
FILING DATE: 03-MAY-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/565,544  
FILING DATE: 10-AUG-1990  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/084,653

US-08-049-783-1/C  
Sequence 1, Application US/08049783  
Patent No. 5439881  
GENERAL INFORMATION:  
APPLICANT: Narva, Kenneth E  
APPLICANT: Schwab, George E  
APPLICANT: Payne, Jewel M  
TITLE OF INVENTION: Gene Encoding No. 5439881e1 Nematode-Active  
TITLE OF INVENTION: Toxins Cloned from Bacillus thuringiensis Isolates  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jeff Lloyd  
STREET: 2421 N.W. 41st Street  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/049,783  
FILING DATE: 19930419  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lloyd, Jeff  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 1 (PS33F2):  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3771 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
INDIVIDUAL ISOLATE: 33f2  
IMMEDIATE SOURCE:  
CLONE: 33f2a  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: 4..24  
OTHER INFORMATION: /function="oligonucleotide  
OTHER INFORMATION: /hybridization probe"  
OTHER INFORMATION: /product="GCA/T ACA/T TTA AAT GAA GTA/T TAT"  
OTHER INFORMATION: /standard\_name="probe a"  
OTHER INFORMATION: /note="probe A"  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: 13..33  
OTHER INFORMATION: /function="oligonucleotide  
OTHER INFORMATION: /hybridization probe"  
OTHER INFORMATION: /product="AAT GAA GTA/T TAT CCA/T GTA/T AAT"  
OTHER INFORMATION: /standard\_name="probe B"  
OTHER INFORMATION: /label="probe-b"  
OTHER INFORMATION: /note="probe b"  
US-08-049-783-1  
Query Match 72.4%; Score 15.2; DB 1; Length 3771;  
Best Local Similarity 85.0%; Pred. No. 2.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
DB 1078 TAGGCGATTGGAGATGTG 20  
1 TAGGCGATTGGAGATGTG 20  
1078 TAGGCGATTGGAGATGTG 1059

RESULT 10  
US-08-158-232-5/C  
Sequence 5, Application US/08158232  
Patent No. 5596071  
GENERAL INFORMATION:  
APPLICANT: Payne, Jewel  
APPLICANT: Kennedy, M. Keith  
APPLICANT: Randall, John Brooks  
APPLICANT: Meier, Henry  
APPLICANT: Vick, Heidi Jane  
APPLICANT: Fonceterada, Luis  
APPLICANT: Schaeff, H. Ernest  
APPLICANT: Schwab, George E.  
APPLICANT: Fu, Jenny  
TITLE OF INVENTION: No. 5596071e1 Bacillus thuringiensis Toxins Active  
TITLE OF INVENTION: Against Hymenopteran Pests  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/158,232  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/887,980  
FILING DATE: 22-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/797,645  
FILING DATE: 25-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/703,977  
FILING DATE: 22-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/SCI104.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3771 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
INDIVIDUAL ISOLATE: 33f2  
IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMYC2316) B-18785  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: 4..24  
OTHER INFORMATION: /function="oligonucleotide  
OTHER INFORMATION: /hybridization probe"  
OTHER INFORMATION: /product="GCA/T ACA/T TTA AAT GAA GTA/T TAT"  
OTHER INFORMATION: /standard\_name="probe a"  
OTHER INFORMATION: /note="probe A"  
FEATURE:  
NAME/KEY: misc-feature

RESULT 6  
US-09-262-773-5  
; Sequence 5, Application US/09262773  
; Patent No. 6225451  
; GENERAL INFORMATION:  
; APPLICANT: Ballinger, Dennis G.  
; APPLICANT: Ding, Wei  
; APPLICANT: Wagner, Susanne  
; APPLICANT: Hess, Mark A.  
; TITLE OF INVENTION: SUSCEPTIBILITY GENE CHD1  
; FILE REFERENCE: Myriad 3  
; CURRENT APPLICATION NUMBER: US/09/262,773  
; CURRENT FILING DATE: 1999-03-04  
; NUMBER OF SEQ ID NOS: 210  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 3264  
; TYPE: DNA  
; ORGANISM: human  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (98)..(496)  
US-09-262-773-5

Query Match 72.4%; Score 15.2; DB 4; Length 3264;  
Best Local Similarity 85.0%; Pred. No. 2.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGGGATTGGAGATTGCA 21  
DB 2993 AGGGGATTGGAACTGTCA 3012

RESULT 7  
US-09-262-773-1  
; Sequence 1, Application US/09262773  
; Patent No. 6225451  
; GENERAL INFORMATION:  
; APPLICANT: Ballinger, Dennis G.  
; APPLICANT: Ding, Wei  
; APPLICANT: Wagner, Susanne  
; APPLICANT: Hess, Mark A.  
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE  
; FILE REFERENCE: Myriad 3  
; CURRENT APPLICATION NUMBER: US/09/262,773  
; CURRENT FILING DATE: 1999-03-04  
; NUMBER OF SEQ ID NOS: 210  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 3268  
; TYPE: DNA  
; ORGANISM: human  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (98)..(2041)  
US-09-262-773-1

Query Match 72.4%; Score 15.2; DB 4; Length 3268;  
Best Local Similarity 85.0%; Pred. No. 2.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGGGATTGGAGATTGCA 21  
DB 2997 AGGGGATTGGAACTGTCA 3016

RESULT 8  
US-07-876-280-5/c  
; Sequence 5, Application US/07876280  
; Patent No. 5262158

GENERAL INFORMATION:  
; APPLICANT: Payne, Jewel M.  
; APPLICANT: Cannon, Raymond J.C.  
; APPLICANT: Bagley, Angela L.  
; TITLE OF INVENTION: No. 5262158el Bacillus thuringiensis Isolates for  
; TITLE OF INVENTION: Controlling Acarides  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David R. Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/876,280  
; FILING DATE: 19920430  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: M/S 104  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3771 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Bacillus thuringiensis  
; INDIVIDUAL ISOLATE: 31f2  
; IMMEDIATE SOURCE:  
; CLONE: E. coli NM522(pMYC 2316) B-18785  
FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 4..24  
; OTHER INFORMATION: /function= "oligonucleotide"  
; OTHER INFORMATION: /product= "hybridization probe"  
; OTHER INFORMATION: /product= "GCA/T ACA/T TTA AAT GAA GTA/T TAT"  
; OTHER INFORMATION: /standard\_name= "probe a"  
; OTHER INFORMATION: /note= "probe A"  
FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 13..33  
; OTHER INFORMATION: /function= "oligonucleotide"  
; OTHER INFORMATION: /product= "hybridization probe"  
; OTHER INFORMATION: /product= "AAT GAA GTA/T TAT CCA/T GTA/T AAT"  
; OTHER INFORMATION: /label= "probe-b"  
; OTHER INFORMATION: /note= "probe b"  
US-07-876-280-5

Query Match 72.4%; Score 15.2; DB 1; Length 3771;  
Best Local Similarity 85.0%; Pred. No. 2.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGCGGATTGGAGATTGTG 20  
DB 1078 TAGCGGATTGTAGATTGTG 1059

RESULT 9



APPLICANT: Algate, Paul A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
FILE REFERENCE: 210121.462C2  
CURRENT APPLICATION NUMBER: US/09/404,879A  
CURRENT FILING DATE: 1999-09-24  
NUMBER OF SEQ ID NOS: 393  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 46  
LENGTH: 481  
TYPE: DNA  
ORGANISM: Homo sapien  
US-09-404-879A-46

Query Match 72.4%; Score 15.2; DB 4; Length 481;  
Best Local Similarity 85.0%; Pred. No. 1.8e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGATTGGAGATTGTG 20  
Db 117 TAGGCTATTGGAGATGCTG 98

RESULT 3  
US-08-993-228-20  
Sequence 20, Application US/08993228  
Patent No. 5976838  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: Lavallee, Edward R.  
APPLICANT: Racie, Lisa A.  
APPLICANT: Treacy, Maurice  
APPLICANT: Spaulding, Vikki  
APPLICANT: Agostino, Michael J.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
FILE REFERENCE: MYriad 3  
NUMBER OF SEQUENCES: 33  
TITLE OF INVENTION: ENCODING THEM  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/993,228  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sprunger, Suzanne A.  
REGISTRATION NUMBER: 41,323  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8284  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3113 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-993-228-20

Query Match 72.4%; Score 15.2; DB 2; Length 3113;  
Best Local Similarity 85.0%; Pred. No. 2.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGGATTGGAGATTGTGA 21  
Db 407 AGGGAAATTGGAGGCTGTGA 426

RESULT 4  
US-09-262-773-7  
Sequence 7, Application US/09262773  
Patent No. 6225451  
GENERAL INFORMATION:  
APPLICANT: Ballinger, Dennis G.  
APPLICANT: Ding, Wei  
APPLICANT: Wagner, Susanne  
APPLICANT: Hess, Mark A.  
TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE  
FILE REFERENCE: MYriad 3  
CURRENT APPLICATION NUMBER: US/09/262,773  
CURRENT FILING DATE: 1999-03-04  
NUMBER OF SEQ ID NOS: 210  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 7  
LENGTH: 3240  
TYPE: DNA  
ORGANISM: human  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (98)..(523)  
US-09-262-773-7

Query Match 72.4%; Score 15.2; DB 4; Length 3240;  
Best Local Similarity 85.0%; Pred. No. 2.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGGATTGGAGATTGTGA 21  
Db 2973 AGGGATTGGAACTGTGA 2992

RESULT 5  
US-09-262-773-3  
Sequence 3, Application US/09262773  
Patent No. 6225451  
GENERAL INFORMATION:  
APPLICANT: Ballinger, Dennis G.  
APPLICANT: Ding, Wei  
APPLICANT: Wagner, Susanne  
APPLICANT: Hess, Mark A.  
TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE  
FILE REFERENCE: MYriad 3  
CURRENT APPLICATION NUMBER: US/09/262,773  
CURRENT FILING DATE: 1999-03-04  
NUMBER OF SEQ ID NOS: 210  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3  
LENGTH: 3244  
TYPE: DNA  
ORGANISM: human  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (98)..(2017)  
US-09-262-773-3

Query Match 72.4%; Score 15.2; DB 4; Length 3244;  
Best Local Similarity 85.0%; Pred. No. 2.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:45:20 ; Search time 5.03708 Seconds

(without alignments)  
1278.561 Million cell updates/sec

Title: US-09-477-082-31

Perfect score: 21

Sequence: 1 taaggagattggagattgtga 21

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA:\*

1: /cgn2\_6/ptodata/1/lna/5A.COMB.seq:\*

2: /cgn2\_6/ptodata/1/lna/5B.COMB.seq:\*

3: /cgn2\_6/ptodata/1/lna/6A.COMB.seq:\*

4: /cgn2\_6/ptodata/1/lna/6B.COMB.seq:\*

5: /cgn2\_6/ptodata/1/lna/PCRTUS.COMB.seq:\*

6: /cgn2\_6/ptodata/1/lna/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.8	80.0	168575	4	US-09-426-290-1
2	15.2	72.4	481	4	US-09-404-879A-46
3	15.2	72.4	3113	2	US-08-993-228-20
4	15.2	72.4	3240	4	US-09-262-773-7
5	15.2	72.4	3244	4	US-09-262-773-3
6	15.2	72.4	3264	4	US-09-262-773-5
7	15.2	72.4	3268	4	US-09-262-773-1
8	15.2	72.4	3771	1	US-07-876-280-5
9	15.2	72.4	3771	1	US-08-049-783-1
10	15.2	72.4	3771	1	US-08-158-232-5
11	15.2	72.4	3771	1	US-08-304-626-5
12	15.2	72.4	3771	1	US-08-816-301A-5
13	15.2	72.4	3771	1	US-08-611-928-5
14	15.2	72.4	3771	3	US-09-173-891-5
15	15.2	72.4	3771	4	US-09-076-137-5
16	15.2	72.4	3771	5	PCT-US92-03624-5
17	15.2	72.4	5849	4	US-09-134-246-6
18	15.2	72.4	20137	4	US-09-262-773-206
19	15.2	72.4	20138	4	US-09-262-773-9
20	15.2	72.4	23071	4	US-09-262-773-210
21	15.2	72.4	15231	3	US-09-128-155-16
22	15.2	72.4	176373	3	US-09-128-155-17
23	14.8	70.5	22	4	US-09-262-773-187
24	14.8	70.5	543	4	US-09-221-017B-681
25	14.8	70.5	1340	4	US-09-000-062-2
26	14.8	70.5	1340	4	US-09-000-062-4
27	14.8	70.5	1713	4	US-09-000-062-1

28	14.8	70.5	1978	6	5188642-3	Patent No. 5188642
29	14.8	70.5	2643	4	US-09-399-913-56	Sequence 56, Appl
30	14.8	70.5	21040	4	US-08-961-527-55	Sequence 55, Appl
31	14.6	69.5	407	4	US-09-605-785-506	Sequence 506, App
32	14.6	69.5	407	4	US-09-439-313-506	Sequence 506, App
33	14.6	69.5	411	4	US-09-605-785-508	Sequence 508, App
34	14.6	69.5	411	4	US-09-439-313-508	Sequence 508, App
35	14.6	69.5	422	4	US-09-605-785-507	Sequence 507, App
36	14.6	69.5	422	4	US-09-439-313-507	Sequence 507, App
37	14.6	69.5	768	1	US-08-592-126-116	Sequence 116, App
38	14.6	69.5	1693	4	US-09-149-476-224	Sequence 224, Appl
39	14.6	69.5	1753	4	US-09-149-476-56	Sequence 56, Appl
40	14.6	69.5	2380	3	US-08-700-651-4	Sequence 4, Appl
41	14.6	69.5	2887	4	US-08-383-502-14	Sequence 14, Appl
42	14.6	69.5	2887	5	PCT-US96-10521-14	Sequence 14, Appl
43	14.6	69.5	28473	4	US-08-961-527-83	Sequence 83, Appl
44	14.4	68.6	256	2	US-08-365-486A-22	Sequence 22, Appl
45	14.4	68.6	256	4	US-08-880-342-22	Sequence 22, Appl

## ALIGNMENTS

```
RESULT 1
US-09-426-290-1
Sequence 1, Application US/09426290
Patent No. 6410712
GENERAL INFORMATION:
APPLICANT: Berglind Ran Olafsdottir
TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
FILE REFERENCE: 2345.2001-000
CURRENT APPLICATION NUMBER: US/09/426.290
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 1
LENGTH: 168575
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (21181)...(21403)
NAME/KEY: CDS
LOCATION: (95252)...(95430)
NAME/KEY: CDS
LOCATION: (101753)...(101966)
NAME/KEY: CDS
LOCATION: (110324)...(110438)
NAME/KEY: CDS
LOCATION: (124058)...(124278)
NAME/KEY: CDS
LOCATION: (127009)...(127130)
NAME/KEY: CDS
LOCATION: (128910)...(129139)
US-09-426-290-1
Query Match 80.0% Score 16.8 DB 4: Length 168575:
Best Local Similarity 90.0% Pred: 57;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

QY 2 AGGGATTGGAGATTGTGA 21  
||||| ||||| |||||  
DB 29456 AGGGATTGGAGATTGTGA 29475

RESULT 2  
US-09-404-879A-46/c  
Sequence 46, Application US/09404879A  
Patent No. 6468546  
GENERAL INFORMATION:  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: King, Gordon E.

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3974043"
/tissue_type="branchial_arch"
/dev_stage="embryo, 10.5 dpc"
/lab_host="DH10B (phage resistant)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACATCTGACGCGCGCGCGCATGTTTGTGTTTGTGTTTGTGTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."
BASE COUNT      114 a      130 c      191 g      110 t      1 others
ORIGIN

```

```

Query Match      82.9%; Score 17.4; DB 12; Length 546;
Best Local Similarity 94.7%; Pred. No. 1.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      2 AGGGGATTGGAGATTGTG 20
        |||||
Db      477 AGGGGTTGGAGATTGTG 495

```

Search completed: July 8, 2003, 05:51:38  
 Job time : 174.34 secs

RESULT 13  
BE471807 488 bp mRNA linear EST 28-JUL-2000  
LOCUS EST16660 potato stolon, Cornell University Solanum tuberosum cDNA  
DEFINITION BE471807.1 GI:9562298  
ACCESSION BE471807  
VERSION BE471807.1 GI:9562298  
KEYWORDS EST.  
SOURCE potato.  
ORGANISM Solanum tuberosum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
1 (bases 1 to 488)  
van der Hoeven,R., Bezzerides,J., Bachem,C., Horvath,B., Visser,R., Holt,I.E., Liang,F., Hansen,T.S., Uitterlinden,T., Doan,B., Bougri,O., Buell,C.R., Rongning,C.M., Tanksley,S.D. and Baker,B.  
TITLE Generation of ESTs from potato swelling stolons  
JOURNAL Unpublished (1999)  
COMMENT Contact: Research Genetics, Libraries Division  
Tel: 1-800-711-6195  
Email: cdna@resgen.com  
5 prime sequence.

FEATURES  
source  
Location/Qualifiers  
1..488  
/organism="Solanum tuberosum"  
/cultivar="Bintje"  
/db\_xref="taxon:4113"  
/clone="CST830H15"  
/clone\_lib="potato stolon, Cornell University"  
/tissue\_type="axillary buds of stem explants, swelling stolons"  
/dev\_stage="1 to 3 days"  
/lab\_host="SOLR"  
/note="vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; RNA was supplied by Christian Bachem & Beatrix Horvath(Laboratory of Plant Breeding, Dept. of Plant Sciences, Wageningen University, The Netherlands). Total RNA was isolated from developing axillary buds of potato nodal stem cuttings cultured on medium for the introduction of tuber formation as described in Bachem et al. (Plant Journal 1996). Tissue samples were taken of stages corresponding to growing stolons and the early stages of tuber formation."  
BASE COUNT 99 a 108 c 136 g 145 t  
ORIGIN

Query Match 82.9%; Score 17.4; DB 10; Length 488;  
Best Local Similarity 94.7%; Pred. No. 1.3e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGGGATTGGAGATTGTG 20  
|||||  
Db 414 AGGGATTGGAGATTGTG 432

RESULT 14  
BQ118560 535 bp mRNA linear EST 22-JUL-2002  
LOCUS EST604136 mixed potato tissues Solanum tuberosum cDNA clone STMB75  
DEFINITION BQ118560.2 GI:21918968  
ACCESSION BQ118560  
VERSION BQ118560.2 GI:21918968  
KEYWORDS EST.  
SOURCE potato.  
ORGANISM Solanum tuberosum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
1 (bases 1 to 535)  
Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C., Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and

TITLE Karamycheva,S.A.  
JOURNAL Generation of a set of potato cDNA clones for microarray analyses  
COMMENT Unpublished (2002)  
On Apr 17, 2002 this sequence version replaced gi:20170522.  
Contact: Robin Buell  
The Institute for Genomic Research  
9712 Medical Center Dr, Rockville, MD 20850, USA  
Email: potato@tigr.org  
This clone is available through the Research Genetics, contact the Research Genetics for further information 1-800-711-6195 or cdna@resgen.com  
Seq primer: T3.

FEATURES  
source  
Location/Qualifiers  
1..535  
/organism="Solanum tuberosum"  
/cultivar="Kennebec or Bintje"  
/db\_xref="taxon:4113"  
/clone="STMB75"  
/clone\_lib="mixed potato tissues"  
/tissue\_type="mixed tissues"  
/lab\_host="SOLR"  
/note="vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: XhoI; supplier: combination of untreated and phytophthora infestans-treated libraries of stolons, leaves, leaflets, axillary buds of stem explants, petioles, germinating eyes, tubers, or roots."  
BASE COUNT 105 a 121 c 149 g 160 t  
ORIGIN

Query Match 82.9%; Score 17.4; DB 14; Length 535;  
Best Local Similarity 94.7%; Pred. No. 1.4e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGGGATTGGAGATTGTG 20  
|||||  
Db 414 AGGGATTGGAGATTGTG 432

RESULT 15  
BG147712 546 bp mRNA linear EST 01-FEB-2001  
LOCUS mab53d02.x1 Soares\_NMEBA\_branchial\_arch Mus musculus cDNA clone  
DEFINITION IMAGE:3974043 3 similar to TR:Q9V115 Q9V115 CG10286 PROTEIN. ?;  
RNA sequence.  
ACCESSION BG147712  
VERSION BG147712.1 GI:12651120  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 546)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
Tumor Gene Index  
Unpublished (1997)  
Other ESTs: mab53d02.y1  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
cDNA library Preparation: M. Bento Soares, Ph.D., M. Fatima Ronaldo  
Ph.D.  
cDNA library Arranged by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
image.llnl.gov/image/html/resources.shtml  
MG1:1474075  
Possible reversed clone: similarity on wrong strand  
Possible reversed clone: polyt not found  
High quality sequence stop: 488.  
Location/Qualifiers  
1..546



REFERENCE 1 (bases 1 to 665)  
 AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.  
 TITLE Whole genome shotgun sequencing of Brassica oleracea  
 JOURNAL Unpublished (2001)  
 COMMENT Other\_GSSS: BOGGB887F  
 Contact: Chris Town

TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TR  
 Class: sheared ends.

FEATURES  
 source  
 1..665  
 Location/Qualifiers

BASE COUNT 168 a 159 c 139 g 179 t  
 ORIGIN  
 Query Match 84.8%; Score 17.8; DB 17; Length 665;  
 Best Local Similarity 90.5%; Pred. No. 9.5e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATTGTGA 21  
 | ||| ||||| ||||| |||||  
 Db 259 TTGGGGATTGGAGATTGTGA 239

RESULT 9  
 BO998627 710 bp mRNA linear EST 22-AUG-2002  
 LOCUS OG61C03.yg.ab1 OG-EFGHJ lettuce serritola Lactuca sativa cDNA clone  
 DEFINITION OG61C03, mRNA sequence.  
 ACCESSION BO998627  
 VERSION BO998627.1 GI:22433023  
 KEYWORDS EST.  
 SOURCE Lactuca sativa.  
 ORGANISM Lactuca sativa.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;  
 Lactuca.

REFERENCE 1 (bases 1 to 710)  
 AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,  
 Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison  
 P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,  
 Church,S., Jackson,L. and Bradford,K.  
 Lettece and Sunflower ESTs from the Compositae Genome Project  
 http://compgenomics.ucdavis.edu/  
 Unpublished (2002)  
 CONTACT: Alexander Kozik [R.W.Michelmore]  
 Department of Vegetable Crops, R.W.Michelmore Lab  
 University of California at Davis (UCD)  
 Asmundson Hall, UCD, Davis, CA 95616, USA  
 Tel: 1-(530)-742-1742  
 Fax: 1-(530)-752-9659  
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]  
 belongs to contig OG\_CA.Contig1855, see http://cgdb.ucdavis.edu/  
 for details.  
 Plate: OG61 row: c column: 03.

FEATURES  
 source  
 1..710  
 Location/Qualifiers

/organism="Lactuca sativa"  
 /cultivar="L.serritola"  
 /db\_xref="taxon:4236"

/clone="OG61C03"  
 /clone\_lib="OG-EFGHJ lettuce serritola"  
 /lab\_host="E.coli"  
 /note="Vector: pBRCDNA5FIAB: The library was constructed  
 from 10 different sources of RNA from a single genotype.  
 Separate cDNAs were generated using primers that  
 incorporated unique 5' and 3' tags to distinguish each  
 source of RNA. cDNAs were then pooled, size-fractionated,  
 directionally cloned into a custom medium-copy vector and  
 transformations made with four size classes to minimize  
 size bias. Details of each source of RNA and library  
 construction can be obtained at http://cgdb.ucdavis.edu/  
 TAG\_LIB-OG-EFGHJ lettuce serritola  
 TAG\_TISSUE=flowers environmental stress  
 TAG\_SEQ=CGAATCGCGG"

BASE COUNT 174 a 157 c 168 g 211 t  
 ORIGIN  
 Query Match 84.8%; Score 17.8; DB 14; Length 710;  
 Best Local Similarity 90.5%; Pred. No. 9.6e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATTGTGA 21  
 | ||| ||||| ||||| |||||  
 Db 56 TTGGGGATTGGAGATTGTGA 76

RESULT 10  
 BO997349 737 bp mRNA linear EST 22-AUG-2002  
 LOCUS OG61D11.yg.ab1 OG-EFGHJ lettuce serritola Lactuca sativa cDNA clone  
 DEFINITION OG61D11, mRNA sequence.  
 ACCESSION BO997349  
 VERSION BO997349.1 GI:22431745  
 KEYWORDS EST.  
 SOURCE Lactuca sativa.  
 ORGANISM Lactuca sativa.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;  
 Lactuca.

REFERENCE 1 (bases 1 to 737)  
 AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,  
 Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison  
 P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,  
 Church,S., Jackson,L. and Bradford,K.  
 Lettece and Sunflower ESTs from the Compositae Genome Project  
 http://compgenomics.ucdavis.edu/  
 Unpublished (2002)  
 CONTACT: Alexander Kozik [R.W.Michelmore]  
 Department of Vegetable Crops, R.W.Michelmore Lab  
 University of California at Davis (UCD)  
 Asmundson Hall, UCD, Davis, CA 95616, USA  
 Tel: 1-(530)-742-1742  
 Fax: 1-(530)-752-9659  
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]  
 belongs to contig OG\_CA.Contig1855, see http://cgdb.ucdavis.edu/  
 for details.  
 Plate: OG616 row: D column: 11.

FEATURES  
 source  
 1..737  
 Location/Qualifiers

/organism="Lactuca sativa"  
 /cultivar="L.serritola"  
 /db\_xref="taxon:4236"  
 /clone="OG61D11"  
 /clone\_lib="OG-EFGHJ lettuce serritola"  
 /lab\_host="E.coli"  
 /note="Vector: pBRCDNA5FIAB: The library was constructed  
 from 10 different sources of RNA from a single genotype.  
 Separate cDNAs were generated using primers that  
 incorporated unique 5' and 3' tags to distinguish each  
 source of RNA. cDNAs were then pooled, size-fractionated,  
 directionally cloned into a custom medium-copy vector and

QY 1 TAGGGGATTGGAGATTGTGA 21  
 ||||| |||||  
 DB 497 TAGGGGATTGGAGATTGTGA 477

RESULT 6  
 B0002845 607 bp mRNA linear EST 22-AUG-2002  
 LOCUS B0002845  
 DEFINITION OG633P01.Y9.ab1 OG\_EFGHJ lettuce serriola Lactuca sativa cDNA clone  
 OG633P01 mRNA sequence.  
 ACCESSION B0002845  
 VERSION B0002845  
 KEYWORDS GI:22437240  
 SOURCE EST.  
 ORGANISM Lactuca sativa.  
 Lactuca sativa.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;  
 Lactuca.

REFERENCE  
 1 (bases 1 to 607)  
 Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,  
 Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison  
 P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,  
 Church, S., Jackson, L. and Bradford, K.  
 Lettuce and Sunflower ESTs from the Compositae Genome Project  
 http://compgenomics.ucdavis.edu/  
 Unpublished (2002)  
 CONTACT: Alexander Kozik [R.W.Michelmore]  
 Department of Vegetable Crops, R.W.Michelmore Lab  
 University of California at Davis (UCD)  
 Asmundo Hall, UCD, Davis, CA 95616, USA  
 Tel: 1-(530)-742-1742  
 Fax: 1-(530)-752-9659  
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]  
 belongs to contig OG\_CA.Contig1855, see http://cgpdb.ucdavis.edu/  
 for details.  
 Plate: OG633 row: D column: 01.

FEATURES  
 source  
 1. 607  
 Location/Qualifiers  
 /organism="Lactuca sativa"  
 /cultivar="L.seriola"  
 /db\_xref="taxon:4236"  
 /clone="OG633P01"  
 /clone\_1lb="OG\_EFGHJ lettuce serriola"  
 /lab\_host="E.coli"  
 /note="Vector: pBRCDNA51AB; The library was constructed  
 from 10 different sources of RNA from a single genotype.  
 Separate cDNAs were generated using primers that  
 incorporated unique 5' and 3' tags to distinguish each  
 source of RNA. cDNAs were then pooled, size-fractionated,  
 directionally cloned into a custom medium-copy vector and  
 transformations made with four size classes to minimize  
 size bias. Details of each source of RNA and library  
 construction can be obtained at http://cgpdb.ucdavis.edu/  
 TAG\_LIB-OG\_EFGHJ lettuce serriola  
 TAG\_TISSUE=germinating seeds  
 TAG\_SEQ=TCGTGCGCG"

BASE COUNT 145 a 136 c 140 g 186 t

ORIGIN  
 Query Match 84.8%; Score 17.8; DB 14; Length 607;  
 Best Local Similarity 90.5%; Pred. No. 9.3e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATTGTGA 21  
 ||||| |||||  
 DB 56 TTGGCGATTGGAGATTGTGA 76

RESULT 7  
 B0996610 630 bp mRNA linear EST 22-AUG-2002  
 LOCUS B0996610

DEFINITION OG633P01.Y9.ab1 OG\_EFGHJ lettuce serriola Lactuca sativa cDNA clone  
 OG633P01 mRNA sequence.  
 ACCESSION B0996610  
 VERSION B0996610  
 KEYWORDS GI:22431006  
 SOURCE EST.  
 ORGANISM Lactuca sativa.  
 Lactuca sativa.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;  
 Lactuca.

REFERENCE  
 1 (bases 1 to 630)  
 Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,  
 Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison  
 P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,  
 Church, S., Jackson, L. and Bradford, K.  
 Lettuce and Sunflower ESTs from the Compositae Genome Project  
 http://compgenomics.ucdavis.edu/  
 Unpublished (2002)  
 CONTACT: Alexander Kozik [R.W.Michelmore]  
 Department of Vegetable Crops, R.W.Michelmore Lab  
 University of California at Davis (UCD)  
 Asmundo Hall, UCD, Davis, CA 95616, USA  
 Tel: 1-(530)-742-1742  
 Fax: 1-(530)-752-9659  
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]  
 belongs to contig OG\_CA.Contig1855, see http://cgpdb.ucdavis.edu/  
 for details.  
 Plate: OG633 row: F column: 01.

FEATURES  
 source  
 1. 630  
 Location/Qualifiers  
 /organism="Lactuca sativa"  
 /cultivar="L.seriola"  
 /db\_xref="taxon:4236"  
 /clone="OG633P01"  
 /clone\_1lb="OG\_EFGHJ lettuce serriola"  
 /lab\_host="E.coli"  
 /note="Vector: pBRCDNA51AB; The library was constructed  
 from 10 different sources of RNA from a single genotype.  
 Separate cDNAs were generated using primers that  
 incorporated unique 5' and 3' tags to distinguish each  
 source of RNA. cDNAs were then pooled, size-fractionated,  
 directionally cloned into a custom medium-copy vector and  
 transformations made with four size classes to minimize  
 size bias. Details of each source of RNA and library  
 construction can be obtained at http://cgpdb.ucdavis.edu/  
 TAG\_LIB-OG\_EFGHJ lettuce serriola  
 TAG\_TISSUE=germinating seeds  
 TAG\_SEQ=TCGTGCGCG"

BASE COUNT 149 a 140 c 148 g 193 t

ORIGIN  
 Query Match 84.8%; Score 17.8; DB 14; Length 630;  
 Best Local Similarity 90.5%; Pred. No. 9.4e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATTGTGA 21  
 ||||| |||||  
 DB 56 TTGGCGATTGGAGATTGTGA 76

RESULT 8  
 BHS41998 665 bp DNA linear GSS 14-DEC-2001  
 LOCUS BHS41998/c  
 DEFINITION BHS41998 BRASSICA oleracea genomic clone B08GB88, DNA  
 sequence.  
 ACCESSION BHS41998  
 VERSION BHS41998  
 KEYWORDS GI:17793048  
 SOURCE GSS.  
 ORGANISM Brassica oleracea.  
 Brassica oleracea.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;



OY 2 AGCGATTGGAGATTGTGA 21  
 |||||  
 Db 793 AGCGATTGGAGATTGTGA 812

RESULT 4  
 BI958770 715 bp mRNA linear EST 22-OCT-2001  
 LOCUS BI958770/c  
 DEFINITION HVSMN0016116f Hordeum vulgare rachis EST library HVCNDA0015  
 (normal) Hordeum vulgare cDNA clone HVSMN0016116f, mRNA sequence.  
 ACCESSION BI958770  
 VERSION BI958770.1 GI:16310025  
 KEYWORDS EST  
 SOURCE Hordeum vulgare  
 ORGANISM Hordeum vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae

REFERENCE  
 1 (bases 1 to 715)  
 Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Chin, A., Begum, D.,  
 Frisch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons  
 J., Oates, R. and Main, D.

TITLE Development of a genetically and physically anchored EST resource  
 for barley genomics: Morex rachis cDNA library  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: twing@clemson.edu  
 Total hg bases = 214  
 Seq primer: AATTAACCTCAGTAAGCG  
 High quality sequence start: 14  
 High quality sequence stop: 379.  
 Location/Qualifiers

FEATURES  
 SOURCE  
 1..715  
 /organism="Hordeum vulgare"  
 /cultivar="Morex"  
 /db\_xref="taxon:4513"  
 /clone="HVSMN0016116f"  
 /clone\_1lb="Hordeum vulgare rachis EST library HVCNDA0015  
 (normal)"  
 /tissue="rachis"  
 /lab\_host="TUC121"  
 /note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
 XhoI; Plants were grown at Washington State University,  
 Pullman, WA in a greenhouse, the rachises were excised and  
 frozen in liquid nitrogen (Kleinhofs lab). In the TJ Close  
 lab at the University of California, Riverside total RNA  
 was prepared, poly(A) was purified, one primary  
 unamplified cDNA library was made, and 1 million pfu were  
 in vivo excised to give pBluescript SK(-) cDNA phagemids  
 (Chin). Phagemids were plated and picked at the Clemson  
 University Genomics Institute (CUGI) (Begum, Palmer,  
 Frisch, Atkins and Wing). Plasmid DNA preparations, DNA  
 sequencing and sequence analysis were performed at CUGI  
 (Wing, Yu, Frisch, Henry, Simmons, Rambo, Main). The  
 sequence has been trimmed to remove vector sequence and  
 contains a minimum of 100 bases of phred value 20 or  
 above. For more details on library preparation and  
 sequence analysis see  
 http://www.genome.clemson.edu/projects/barley. To order  
 this clone see http://www.genome.clemson.edu/orders/Also  
 see Close TJ, Wing R, Kleinhofs A, Wise R (2001)  
 Genetically and physically anchored EST resources for  
 barley genomics. Barley Genetics Newsletter 31:29-30.  
 (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

BASE COUNT 155 a 172 c 175 g 187 t 26 others  
 Query Match 85.7%; Score 18; DB 13; Length 715;

Best Local Similarity 100.0%; Pred. No. 7.9e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 4 GCGATTGGAGATTGTGA 21  
 |||||  
 Db 356 GCGATTGGAGATTGTGA 339

RESULT 5  
 A0862416/c 512 bp DNA linear GSS 03-NOV-1999  
 LOCUS A0862416/c  
 DEFINITION nbe00018H14f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic  
 clone nbe00018H14f, DNA sequence.  
 ACCESSION A0862416  
 VERSION A0862416.1 GI:6212873  
 KEYWORDS GSS.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE  
 1 (bases 1 to 512)  
 Wing, R.A. and Dean, R.A.  
 A BAC End Sequencing Framework to Sequence the Rice Genome  
 JOURNAL Unpublished (1998)  
 COMMENT Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: twing@clemson.edu  
 Seq primer: TAATACGACTCAGTAAGCG  
 Class: BAC ends  
 High quality sequence start: 46  
 High quality sequence stop: 449.  
 Location/Qualifiers

FEATURES  
 SOURCE  
 1..512  
 /organism="Oryza sativa"  
 /strain="Japonica"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:4530"  
 /clone="nbe00018H14f"  
 /clone\_1lb="CUGI Rice BAC Library (EcoRI)"  
 /tissue="leaf"  
 /lab\_host="E. coli DH10B"  
 /note="Vector: pBACindigo; Site\_1: EcoRI; Site\_2: EcoRI;  
 Rice is the most important food crop in the world. Half of  
 the world population, especially those inhabiting highly  
 populated areas of the humid tropics and subtropics, rely  
 on rice as their primary source of carbohydrate. with a  
 monocotyledonous rice is a diploid plant (2n=24) with a  
 haploid genome equivalent of 431 Mbp (Arumuganathan and  
 Earle, 1991). The relatively small genome of rice, three  
 times larger than that of Arabidopsis, makes it suitable  
 for genomic studies. In order to facilitate positional  
 cloning, physical mapping and genome sequencing of rice,  
 we have constructed a BAC library from Oryza sativa.  
 Nipponbare variety using EcoRI as the cloning enzyme. The  
 library contains 55,296 clones with an average insert size  
 of 121 kb providing approximately 15 haploid genome  
 equivalents. The deep coverage allows the isolation a  
 particular sequence with a probability of 99.9%. Three  
 high density filters, each containing 18,432 clones  
 (doubly spotted), represent the whole library for colony  
 screening and can be requested from the Clemson University  
 BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT 174 a 101 c 81 g 154 t 2 others  
 ORIGIN  
 Query Match 84.8%; Score 17.8; DB 17; Length 512;  
 Best Local Similarity 90.5%; Pred. No. 9.1e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

High quality sequence stop: 527.

FEATURES  
Location/Qualifiers  
1..527

Source  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/clone="MPMGp637-13F2:MPMGp637F0213"  
/clone\_id="zebrafish shield stage whole embryo CDNA library MPMGP637"  
/tissue\_type="whole embryo"  
/dev\_stage="shield stage, 6 hrs post-fertilisation"  
/lab\_host="E.coli, XL1 blue MRF"  
/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; oligo-dT-NotI primed, SalI adaptors, directionally cloned, library normalised by oligonucleotide fingerprinting"  
BASE COUNT 155 a 120 c 143 g 108 t 1 others  
ORIGIN

Query Match 87.6%; Score 18.4; DB 13; Length 527;  
Best Local Similarity 95.0%; Pred. No. 5e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGGGGATTGGAGATTGTGA 21  
|||||  
Db 472 AGGTGATTGGAGATTGTGA 491

RESULT 2 706 bp DNA linear GSS 05-OCT-2000  
A2507082  
LOCUS IM0348K01R mouse 10kb plasmid UUGC1M library Mus musculus genomic  
DEFINITION  
ACCESSION A2507082  
VERSION A2507082.1 GI:10688398  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS

1 (bases 1 to 706)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rilly,  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.  
and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)

CONTACT: Robert B. Weiss  
University of Utah Genome Center  
University of Utah

Rm 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA

Tel: 801 585 5606  
Fax: 801 585 7177

Email: dunn@genetics.utah.edu  
Insert length: 10000 Std Error: 0.00  
Plate: 0348 row: K column: 01

Seq primer: CACACAGAAACACACTATGAC  
Class: plasmid ends

High quality sequence stop: 706.

Location/Qualifiers  
1..706

Source  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0348K01"  
/clone\_id="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. coli strain XL10-Gold, TI-resistant, F-"  
/note="Vector: pMD29v. Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptor DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of pMD42 (G1473211415B1AF19072.1) a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptor mouse DNA was annealed to  
adaptor vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 209 a 150 c 163 g 184 t  
ORIGIN

Query Match 87.6%; Score 18.4; DB 17; Length 706;  
Best Local Similarity 95.0%; Pred. No. 5.2e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGGGGATTGGAGATTGTGA 21  
|||||  
Db 445 AGGGGATTGGAGATTGTGA 464

RESULT 3 939 bp mRNA linear EST 16-AUG-2002  
B0899695  
LOCUS AGENCOURT.8750911 NIH\_MGC\_130 Mus musculus CDNA clone IMAGE:6336113  
DEFINITION  
ACCESSION B0899695  
VERSION B0899695.1 GI:22291697  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS

1 (bases 1 to 939)  
NIH-NGC http://imgc.ncl.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

JOURNAL

CONTACT: Robert Strausberg, Ph.D.  
Email: cga@bbs-rcmail.nih.gov  
Tissue Procurement: Mark Macconochie, Ph.D. and Nancy L. Freeman,  
Ph.D.

CDNA Library Preparation: Resgen, Invitrogen Corp.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation  
(LLNL)

Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

Plate: LLAM13799 row: J column: 18  
High quality sequence stop: 634.

Location/Qualifiers  
1..939

Source  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6336113"  
/clone\_id="NIH\_MGC\_130"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: otcysts; Vector: pCMV-SPORT6.1.cdb;  
Site\_1: BclRV; Site\_2: NotI; Cloned unidirectionally.  
Primer: Oligo dT. Average insert size 1.95 kb.  
Constructed by Resgen, Invitrogen Corp. Note: this is a  
NIH\_MGC Library."  
BASE COUNT 250 a 197 c 281 g 210 t 1 others  
ORIGIN  
Query Match 87.6%; Score 18.4; DB 14; Length 939;  
Best Local Similarity 95.0%; Pred. No. 5.5e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:04:24 ; Search time 158.34 Seconds  
(without alignments)  
2147.949 Million cell updates/sec

Title: US-09-477-082-31

Perfect score: 21  
Sequence: 1 tagggattgagattgtga 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
EST.\*  
1: em\_estda:\*  
2: em\_esthm:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hnc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hnc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estcom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_liv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vtl:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	87.6	527	13	B1891021
2	18.4	87.6	706	17	AZ507082
3	18.4	87.6	938	14	BQ889855
4	18.4	87.6	715	13	B1988770
5	17.8	84.8	512	17	AQ862416
6	17.8	84.8	607	14	BQ002845

Result No.	Score	Query Match	Length	DB ID	Description
7	17.8	84.8	630	14	BQ966610
8	17.8	84.8	665	17	BH541998
9	17.8	84.8	710	14	BQ986827
10	17.8	84.8	737	14	BQ997349
11	17.8	84.8	742	14	BQ859778
12	17.8	84.8	748	14	BQ004608
13	17.4	82.9	488	10	BE471807
14	17.4	82.9	535	14	BQ118560
15	17.4	82.9	546	12	BG147712
16	17.4	82.9	554	10	AW735936
17	17.4	82.9	566	13	B1479559
18	17.4	82.9	621	12	BG597208
19	17.4	82.9	649	13	BQ078834
20	17.4	82.9	685	17	BQ140974
21	17.4	82.9	688	14	BQ027185
22	17.4	82.9	696	14	BQ045836
23	17.4	82.9	714	17	AG092683
24	17.4	82.9	737	13	BK113418
25	17.4	82.9	743	12	BG597487
26	17.4	82.9	767	14	BQ113951
27	17.4	82.9	1101	17	CNS00KAG
28	17.4	82.9	1151	14	BW914796
29	17.4	82.9	663	13	B1271871
30	17.4	82.9	706	13	B1958820
31	17.4	82.9	1101	17	CNS00G9U
32	16.8	80.0	262	12	BF772261
33	16.8	80.0	292	9	AV166017
34	16.8	80.0	351	13	B1051652
35	16.8	80.0	353	17	AQ045614
36	16.8	80.0	367	10	AM315961
37	16.8	80.0	374	12	BE976571
38	16.8	80.0	396	9	A1507111
39	16.8	80.0	397	17	BH529081
40	16.8	80.0	407	9	A1736949
41	16.8	80.0	413	12	BF993537
42	16.8	80.0	438	17	BH656193
43	16.8	80.0	441	9	A1998088
44	16.8	80.0	443	17	AQ216417
45	16.8	80.0	444	17	B72074

## ALIGNMENTS

RESULT 1  
LOCUS B1891021  
DEFINITION ZF637-3-000276 zebrafish shield stage whole embryo cDNA library  
ACCESSION B1891021  
VERSION B1891021.1 GI:16098292  
KEYWORDS mRNA sequence.  
SOURCE MPMGP637 Danio rerio cDNA clone MPMGP637\_F0213 5',  
ORGANISM zebrafish.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
Cyprinidae; Danio.  
REFERENCE 1 (bases 1 to 527)  
AUTHORS Clark, M., Aanstad, P., Hennig, S., Johnson, S.L. and Lehrach, H.  
TITLE EST sequencing of a zebrafish shield stage cDNA library normalised  
by oligonucleotide fingerprinting  
JOURNAL Unpublished (2001)  
COMMENT Contact: Hennig S  
Laboraty 123, dept. Lehrach  
Max-Planck-Institut fuer Molekulare Genetik  
Inhestr. 63-73, D-14195 Berlin, Germany  
Tel: +49 30 8413 1612  
Fax: +49 30 8413 1380  
Email: hennig@molgen.mpg.de  
5' EST sequencing of clones from a zebrafish shield stage library,  
normalised from 55,000 starting clones by oligonucleotide  
fingerprinting

CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB57737-AB572072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 1199 BP; 335 A; 360 C; 320 G; 184 T; 0 other;

Query Match 80.0%; Score 16.8; DB 23; Length 1199;

Best Local Similarity 90.0%; Pred. No. 2.3e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGATTGGAGATTGTG 20  
 | ||||| ||||| |||||

DB 1012 TGGGGATGTGAGATTGTG 993

RESULT 15

AAA97385/c  
 ID AAA97385 standard; DNA; 2325 BP.

AC AAA97385;

DT 29-JAN-2001 (first entry)

DE Pea p1a2 gene light-repressible promoter.

KW GTP-binding protein p1a2; pea; light-repressible promoter;

KW photoinhibitory; expression cassette; transgenic plant;

KW deterioration prevention; storage; ds.

OS Pisum sativum.

PN WO200055313-A1.

PD 21-SEP-2000.

PF 03-MAR-2000; 2000WO-JP01269.

PR 12-MAR-1999; 99JP-0066551.

PA (SUNR) SUNTORY LTD.

PI Sasaki Y, Nagano Y, Inaba T;

PI WPI; 2000-587526/55.

PT New DNA fragment or promoter for expressing a target gene, specifically  
 PT under photoinhibitory conditions, and for transforming a plant cell or  
 PT plant to improve quality and prevent deterioration during storage

PS Claim 3; Page 34-35; 49pp; Japanese.

CC The invention relates to a light-repressible promoter (AAA97385), or  
 CC active fragments thereof (AAA97383, AAA97384), from the gene encoding  
 CC the pea GTP-binding protein p1a2. The invention also relates to an  
 CC expression cassette containing the p1a2 promoter or its active  
 CC fragments for the expression of a gene under photoinhibitory or dark  
 CC conditions in a plant, and to transgenic plants, their descendants  
 CC and plant tissues comprising the expression cassette. The expression  
 CC cassette of the invention can be used to generate transgenic plants in  
 CC which deterioration during storage in the dark is prevented. This is  
 CC particularly useful for agricultural products. The present sequence  
 CC represents the pea p1a2 promoter.

SO Sequence 2325 BP; 791 A; 451 C; 317 G; 766 T; 0 other;

Query Match 80.0%; Score 16.8; DB 21; Length 2325;

Best Local Similarity 90.0%; Pred. No. 2.4e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGGGATTGGAGATTGTGA 21  
 | ||||| ||||| |||||  
 DB 2027 AAGGGATTGGAAATTGTGA 2008

Search completed: July 8, 2003, 03:11:16  
 Job time: 21.9513 secs



PR With gene regulation, useful for analysing cytosine methylations for  
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency  
PR disease  
PS Claim 1: SEQ ID No 214; 26pp; English.  
XX  
CC The invention relates to 224 nucleic acid sequences comprising at least  
CC 18 bases of a chemically pretreated gene associated with gene regulation  
CC selected from 43 known genes (or complementary sequences). The  
CC chemical pretreatment converts cytosine bases unmethylated at the  
CC 5-position to uracil or another base with hybridisation behaviour  
CC dissimilar to cytosine, to enable analysis of cytosine methylations.  
CC The DNA sequences, oligomers (or sets/arrays) and method are  
CC useful in the diagnosis of diseases (or predisposition to diseases)  
CC associated with gene regulation and in therapy of such diseases, by  
CC enabling analysis of the cytosine methylation patterns of such genes,  
CC kits are provided. They are especially useful in diagnosis  
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac  
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,  
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,  
CC preclampsia, graft versus-host disease. The present sequence is a  
CC sequence included in the sequence data for this specification and is  
CC associated with the human gene regulation-associated genes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pcl\_sequences  
XX  
SQ Sequence 16914 BP; 4573 A; 449 C; 4019 G; 7823 T; 50 other;  
Query Match 81.0%; Score 17; DB 24; Length 16914;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TAGGGATTGGAGATT 17  
DB 5123 TAGGGATTGGAGATT 5139  
RESULT 13  
AAC52475  
ID AAC52475 standard; DNA; 898 BP.  
AC AAC52475;  
XX  
DX 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 71461.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
XX Arabidopsis thaliana.  
OS  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-01231825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134258.  
PR 14-MAY-1999; 99US-0134259.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138040.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139839.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141847.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142056.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143547.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0143625.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.

XX OS Homo sapiens.  
 XX PN WO200200928-A2.  
 XX PD 03-JAN-2002.  
 XX PF 02-JUL-2001; 2001WO-EP07537.  
 XX PR 30-JUN-2000; 2000DE-1032529.  
 XX PR 01-SEP-2000; 2000DE-1043826.  
 XX PA (EPIG-) EPIGENOMICS AG.  
 XX PI Olek A, Piepenbrock C, Berlin K;  
 XX DR WPI; 2002-110909/17.  
 XX PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 XX PS Claim 1; SEQ ID NO 1435; 32pp + Sequence Listing; German.  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 XX SQ Sequence 13377 BP; 3915 A; 146 C; 3046 G; 6270 T; 0 other;  
 Query Match 84.8%; Score 17.8; DB 24; Length 13377;  
 Best Local Similarity 90.5%; Pred. No. 98;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 TAGGGGATTGGAGATTGCA 21  
 Db 1704 TAGGGATTGGAGATTGTTA 1724  
 RESULT 11  
 ABL70316  
 ID ABL70316 standard; DNA; 16914 BP.  
 AC ABL70316;  
 XX 01-JUL-2002 (first entry)  
 DE Chemically treated cell signalling DNA sequence complementary to #103.  
 XX Cell signalling; cytosine methylation; cell signalling disease;  
 KW cancer; tumour; cytostatic; ds.  
 XX OS Unidentified.  
 XX PN WO200202807-A2.  
 XX PD 10-JAN-2002.  
 XX PF 29-JUN-2001; 2001WO-EP07471.  
 XX PR 30-JUN-2000; 2000DE-1032529.  
 XX PR 01-SEP-2000; 2000DE-1043826.  
 XX PA (EPIG-) EPIGENOMICS AG.  
 XX PI Olek A, Piepenbrock C, Berlin K;  
 XX DR WPI; 2002-154758/20.

XX OS Homo sapiens.  
 XX PN WO200200928-A2.  
 XX PD 03-JAN-2002.  
 XX PF 02-JUL-2001; 2001WO-EP07537.  
 XX PR 30-JUN-2000; 2000DE-1032529.  
 XX PR 01-SEP-2000; 2000DE-1043826.  
 XX PA (EPIG-) EPIGENOMICS AG.  
 XX PI Olek A, Piepenbrock C, Berlin K;  
 XX DR WPI; 2002-110909/17.  
 XX PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 XX PS Claim 1; SEQ ID NO 206; 24pp+sequence listing; English.  
 CC The invention relates to a nucleic acid comprising a sequence of at least  
 CC 18 bases of a segment of chemically pre-treated DNA of genes associated  
 CC with cell signalling. The activity of the modified sequences of the  
 CC invention may be described as cytostatic. The object of the invention is  
 CC to provide the chemically modified DNA of genes associated with cell  
 CC signalling, as well as oligonucleotides and/or PNA-oligomers for  
 CC detecting cytosine methylations, as well as a method which is  
 CC particularly suitable for the diagnosis and/or therapy of genetic and  
 CC epigenetic parameters of genes associated with cell signalling. The  
 CC chemically modified DNA provided by the invention is useful for diagnosis  
 CC and therapy of diseases such as solid tumours and cancer. The sequences  
 CC given in records ABL70111-ABL70626 represent chemically pre-treated  
 CC genomic DNA's of genes associated with cell signalling.  
 CC Note: The sequence data for this patent is not represented in the printed  
 CC specification, but is based on sequence information supplied by the  
 CC European Patent Office.  
 XX SQ Sequence 16914 BP; 4573 A; 449 C; 4019 G; 7823 T; 50 other;  
 Query Match 81.0%; Score 17; DB 24; Length 16914;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TAGGGGATTGGAGATT 17  
 Db 5123 TAGGGGATTGGAGATT 5139  
 RESULT 12  
 AAS61254  
 ID AAS61254 standard; DNA; 16914 BP.  
 AC AAS61254;  
 XX 29-JAN-2002 (first entry)  
 DE Human gene regulation-associated gene oligonucleotide #209.  
 XX Human; Gene regulation-associated gene; severe combined immunodeficiency;  
 KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;  
 KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;  
 KW renal disease; Preeclampsia; cardiac allograft vascular disease;  
 KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;  
 KW immunostimulant; cardiac; anti-inflammatory; coagulant; antithrombotic;  
 KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.  
 XX OS Homo sapiens.  
 XX PN WO200177375-A2.  
 XX PD 18-OCT-2001.  
 XX PF 06-APR-2001; 2001WO-EP03968.  
 XX PR 06-APR-2000; 2000DE-1019058.  
 XX PR 07-APR-2000; 2000DE-1019173.  
 XX PR 30-JUN-2000; 2000DE-1032529.  
 XX PR 01-SEP-2000; 2000DE-1043826.  
 XX PA (EPIG-) EPIGENOMICS AG.  
 XX PI Olek A, Piepenbrock C, Berlin K;  
 XX DR WPI; 2002-017470/02.  
 XX PT New nucleic acid sequences from chemically modified genes associated

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 33890.  
 XX  
 KW Drosophila: developmental biology; cell signalling; insecticide;  
 KW Pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DX WPI; 2001-656860/75.  
 DR P-PSDB; ABB69033.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 33890; 21pp + Sequence Listing; English.  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB57737-AB872072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 CC Sequence 3536 BP; 1037 A; 820 C; 773 G; 906 T; 0 other;  
 SQ  
 Query Match 84.8%; Score 17.8; DB 23; Length 3536;  
 Best Local Similarity 90.5%; Pred. No. 88;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 TAGGGGATTGGAGATTGTGA 21  
 Db 3107 TTGGGGATTGGAGATTGTGA 3127  
 RESULT 9  
 AAS46475  
 ID AAS46475 standard; DNA; 13377 BP.  
 XX  
 AC AAS46475;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Tumour suppressor gene derived chemically modified sequence #197.  
 XX  
 KW Human: tumour suppressor gene; oncogene; antitumour; cytostatic;  
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
 KW cytosine methylation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200168912-A2.  
 XX  
 PD 20-SEP-2001.  
 XX  
 PF 15-MAR-2001; 2001WO-EP02955.

XX  
 PR 15-MAR-2000; 2000DE-1013847.  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPIDEMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DX WPI; 2001-602752/68.  
 DR  
 XX  
 PT Fragments of chemically modified genes associated with tumour suppressor  
 PT genes and oncogenes, useful in designing primers and probes for  
 PT analysing diseases associated with cytosine methylation state e.g.  
 PT cancer -  
 XX  
 PS Claim 1; SEQ ID NO 197; 27pp; English.  
 XX  
 CC The invention relates to a nucleic acid comprising a sequence of 18  
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
 CC bisulphite, of genes associated with tumour suppression and  
 CC oncogenes having a sequence taken from 536 (actually 533 since  
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
 CC (58) and sequences complementary to (58). The nucleic acid may be a  
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
 CC form part of a set of probes for detecting the cytosine methylation state  
 CC and/or single nucleotide polymorphisms and also to be used in an  
 CC array for analysing diseases associated with CpG dinucleotides e.g.  
 CC cancers and tumours. The probes can also be used in a method for  
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
 CC and/or therapy of existing diseases or the predisposition to specific  
 CC diseases, by analysing cytosine methylations. The parameters may be  
 CC compared to another set of genetic and/or epigenetic parameters, the  
 CC differences serving as basis for diagnosis and/or prognosis events which  
 CC are disadvantageous to patients. The present sequence is one of the  
 CC 533 genomic sequences derived from tumour suppressor genes and  
 CC oncogenes.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 CC Sequence 13377 BP; 3915 A; 146 C; 3046 G; 6270 T; 0 other;  
 SQ  
 Query Match 84.8%; Score 17.8; DB 22; Length 13377;  
 Best Local Similarity 90.5%; Pred. No. 98;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 1 TAGGGGATTGGAGATTGTGA 21  
 Db 1704 TTAGGGATTGGAGATTGTGA 1724  
 RESULT 10  
 ABL33462  
 ID ABL33462 standard; DNA; 13377 BP.  
 XX  
 AC ABL33462;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 1435.  
 XX  
 KW Human: immune system disease; cytosine methylation; antiasthmatic;  
 KW antiatherosclerotic; antianaemic; cytosatic; nootropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; optalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.



CC tumour in which a myc gene is amplified, such as a neuroblastoma.  
 CC Aggressive neuroblastoma, juvenile neuroblastoma (preferred), small-cell  
 CC lung carcinoma, non-small-cell lung carcinoma, colorectal carcinoma, or  
 CC uterine cervical carcinoma can be diagnosed with the new method. A kit  
 CC for screening for a compound that induces death-receptor-mediated  
 CC apoptosis in cells containing an inactivated Casp8 gene is also  
 CC provided.

CC Sequence 21 BP; 5 A; 2 C; 9 G; 5 T; 0 other;

Query Match 84.8%; Score 17.8; DB 21; Length 21;  
 Best Local Similarity 90.5%; Pred. No. 57;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TAGGGATTGGAGATTGTGA 21  
 |||||  
 Db 1 TAGGGATTGGAGATTGTGA 21

RESULT 6  
 ABL22738 ID ABL22738 standard; DNA; 2204 BP.

XX ABL22738;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 19687.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -

PS Claim 1; SEQ ID NO 19687; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB161737-AB16175).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2204 BP; 613 A; 518 C; 507 G; 566 T; 0 other;

Query Match 84.8%; Score 17.8; DB 23; Length 2204;  
 Best Local Similarity 90.5%; Pred. No. 84;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TAGGGATTGGAGATTGTGA 21  
 |||||  
 Db 1558 TTGGGATTGGAGATTGTGA 1578

RESULT 7  
 ABL12786 ID ABL12786 standard; CDNA; 2561 BP.

XX ABL12786;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 32840.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR P-PADB; AB168683.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -

PS Claim 1; SEQ ID NO 32840; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB161737-AB16175).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2561 BP; 750 A; 531 C; 533 G; 747 T; 0 other;

Query Match 84.8%; Score 17.8; DB 23; Length 2561;  
 Best Local Similarity 90.5%; Pred. No. 85;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TAGGGATTGGAGATTGTGA 21  
 |||||  
 Db 225 TTGGGATTGGAGATTGTGA 245

RESULT 8

AB113136 ID ABL13136 standard; CDNA; 3536 BP.

XX ABL13136;

DT 26-MAR-2002 (first entry)

CC the specification such as ACCPN, ADFN, or AFDI and comprising one of 350  
 CC sequences (ABN79984-ABN80333) or their complements. The invention is  
 CC useful for the diagnosis or therapy of diseases associated with  
 CC development genes, in particular disease related to homeobox containing  
 CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes  
 CC associated with congenital heart disease, epilepsy, diseases related to  
 CC histone deacetylation, Curraño syndrome, diseases related with the  
 CC development of the brain and limb girdle muscular dystrophy and dwarfism.  
 CC Oligomers specific to each of the genes are useful for detecting the  
 CC methylation state of all CpG dinucleotides within the 350 sequences or  
 CC (II) and their complementary sequences, as primer oligonucleotides for  
 CC the amplification of the 350 sequences, (II) and/or their complements and  
 CC as oligomer probes for detecting the cytosine methylation state and/or  
 CC single nucleotide polymorphisms (SNPs).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification but is based on sequence information supplied to Derwent by  
 CC the European Patent Office.

XX SQ Sequence 5518 BP; 1383 A; 1452 C; 1452 G; 2541 T; 0 other;

Query Match 87.6%; Score 18.4; DB 24; Length 5518;  
 Best Local Similarity 95.0%; Pred. No. 48;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAGGGGATTGGAGATTGTG 20  
 |||||

DB 46 TAGGGGATTGGAGATTGTG 65

RESULT 4  
 ABL32217  
 ID ABL32217 standard; DNA; 5518 BP.

AC ABL32217;  
 XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 190.

XX Human: immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antihemetic; antirheumatic; antidiabetic; antipsoriatic;  
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043626.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation  
 XX Claim 1; SEQ ID NO 190; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences

CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.

XX SQ Sequence 5518 BP; 1383 A; 1452 C; 1452 G; 2541 T; 0 other;

Query Match 87.6%; Score 18.4; DB 24; Length 5518;  
 Best Local Similarity 95.0%; Pred. No. 48;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAGGGGATTGGAGATTGTG 20  
 |||||

DB 46 TAGGGGATTGGAGATTGTG 65

RESULT 5

AA51818  
 ID AA51818 standard; DNA; 21 BP.

XX AA51818;

XX 31-OCT-2000 (first entry)

DE Sense primer for CASP8 methylated-specific bisulfite treated DNA.

XX CASP8; caspase-8; promoter; inactivation; methylation; cysteine protease;  
 XX tumour suppressor; chromosome 2q33-34; neuroblastoma; cancer;

XX death receptor; apoptosis; cytosolic; gene therapy; primer; ss.

XX Homo sapiens.

XX WO200039347-A1.

XX 06-JUL-2000.

XX 30-DEC-1999; 99WO-US31280.

XX 31-DEC-1998; 98US-0114308.

XX (SUJD-) ST JUDE CHILDREN'S RES HOSPITAL.

XX Kidd VJ, Lahti JM, Teitz T;

XX WPI; 2000-452423/39.

XX Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or  
 PT prognosing cancer, comprises detecting a modification of genomic DNA  
 PT comprising the CASP8 gene that results in inactivation of the gene  
 XX Example 3; Page 61; 107pp; English.

XX Investigation of silencing of caspase-8 (CASP8) gene by methylation was  
 CC analyzed using methylation-sensitive PCR analysis. Amplification of the  
 CC 5' untranslated region (UTR) of the Casp8 gene was performed in reaction  
 CC mixtures containing bisulfite-treated DNA. Primers AA51818-23 were  
 CC designed to produce a 320 bp fragment in the upstream region of Casp8  
 CC gene extending from nucleotides +221 to +541. Wild type primers were  
 CC used to amplify the corresponding region of untreated genomic DNA.  
 CC Controls without DNA were also performed. CASP8, a cysteine protease, is  
 CC part of the death inducing signaling complex (DISC) associated with the  
 CC Fas receptor. CASP8 is inactivated in cancers, and plays a role of a  
 CC tumour suppressor gene. The CASP8 promoter region sequences, in  
 CC particular Region 1 and Region 2, are crucial to the design and execution  
 CC of the genomic methylation PCR analysis of CASP8 gene inactivation.  
 CC Methylation PCR can be used to examine even minute amounts of patient  
 CC material to demonstrate whether the CASP8 gene expresses an mRNA and  
 CC protein product. The CASP8 gene has been localized to human chromosome  
 CC 2q33-34. The methods are used to diagnose or prognose cancer. Cancer is  
 CC treated by administering a vector that expresses a gene encoding  
 CC functional CASP8 in cells. The cancer that is diagnosed or treated is a

PT comprising the CASP8 gene that results in inactivation of the gene  
 XX  
 PS Example 3; Page 61; 107pp; English.  
 CC Investigation of silencing of caspase-8 (CASP8) gene by methylation was  
 CC analyzed using methylation-sensitive PCR analysis. Amplification of the  
 CC 5' untranslated region (UTR) of the CASP8 gene was performed in reaction  
 CC mixtures containing bisulfite treated DNA. Primers AAS1818-23 were  
 CC designed to produce a 320 bp fragment in the upstream region of CASP8  
 CC gene extending from nucleotides +221 to +541. Wild type primers were  
 CC used to amplify the corresponding region of untreated genomic DNA.  
 CC Controls without DNA were also performed. CASP8, a cysteine protease, is  
 CC part of the death inducing signaling complex (DISC) associated with the  
 CC Fas receptor. CASP8 is inactivated in cancers, and plays a role of a  
 CC tumour suppressor gene. The CASP8 promoter region sequences, in  
 CC particular Region 1 and Region 2, are crucial to the design and execution  
 CC of the genomic methylation PCR analysis of CASP8 gene inactivation.  
 CC Methylation PCR can be used to examine even minute amounts of patient  
 CC material to demonstrate whether the CASP8 gene expresses an mRNA and  
 CC protein product. The CASP8 gene has been localized to human chromosome  
 CC 2q33-34. The methods are used to diagnose or prognose cancer. Cancer is  
 CC treated by administering a vector that expresses a gene encoding  
 CC functional CASP8 in cells. The cancer that is diagnosed or treated is a  
 CC tumour in which a myc gene is amplified, such as a neuroblastoma.  
 CC Aggressive neuroblastoma, juvenile neuroblastoma (prefered), small-cell  
 CC lung carcinoma, non-small-cell lung carcinoma, colorectal carcinoma, or  
 CC uterine cervical carcinoma can be diagnosed with the new method. A kit  
 CC for screening for a compound that induces death-receptor-mediated  
 CC apoptosis in cells containing an inactivated CASP8 gene is also  
 CC provided.  
 CC  
 SQ Sequence 21 BP; 5 A; 0 C; 9 G; 7 T; 0 other;  
 XX  
 Query Match 100.0%; Score 21; DB 21; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 2;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TAGGGGATTGGAGATTGTGA 21  
 DB 1 TAGGGGATTGGAGATTGTGA 21  
 RESULT 2  
 ABL54318  
 ID ABL54318 standard; DNA; 5518 BP.  
 XX  
 AC ABL54318;  
 XX  
 DT 29-JUL-2002 (first entry)  
 XX  
 DE Chemically treated apoptosis gene complementary to gene #9.  
 XX  
 KW Apoptosis; HIV; Bloom syndrome; cardiopathy;  
 KW neurodegenerative disorder; Herpes simplex virus; renal ischaemia;  
 KW amyotrophic lateral sclerosis; cancer; ds.  
 XX  
 OS Unidentified.  
 XX  
 PN WO200177164-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 06-APR-2001; 2001WO-EP03969.  
 XX  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX

DR MPI; 2002-017444/02.  
 XX  
 PT Chemically modified sequences of genes associated with apoptosis are  
 PT useful to determine methylation patterns of genomic DNA samples for  
 PT diagnosis of associated diseases such as cancer  
 XX  
 PS Claim 1; Seq ID #18; 24pp; English.  
 XX  
 CC This invention relates to chemically pre-treated DNA of genes  
 CC associated with apoptosis. The nucleic acids are used to allocate  
 CC patients for specific therapy for HIV infection, Bloom syndrome,  
 CC cardiopathy, aging, neurodegenerative disorders, Herpes simplex virus  
 CC infection, renal ischaemia, amyotrophic lateral sclerosis, solid tumours  
 CC and cancers. This nucleotide sequence represents a chemically  
 CC treated apoptosis gene. Even SEQ ID numbers are the complementary  
 CC DNA strands to the odd SEQ ID numbers. The sequence data for this  
 CC patent is not represented in the printed specification but is based on  
 CC information supplied by the European patent office.  
 CC  
 SQ Sequence 5518 BP; 1383 A; 142 C; 1452 G; 2541 T; 0 other;  
 XX  
 Query Match 87.6%; Score 18.4; DB 24; Length 5518;  
 Best Local Similarity 95.0%; Pred. No. 46;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 TAGGGGATTGGAGATTGTG 20  
 DB 46 TAGGGGATTGGAGATTGTG 65  
 RESULT 3  
 ABLN80019  
 ID ABLN80019 standard; DNA; 5518 BP.  
 XX  
 AC ABLN80019;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Human chemically modified disease associated gene SEQ ID NO 36.  
 XX  
 KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;  
 KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;  
 KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;  
 KW antidiabetic; cytosolic; anticonvulsant; ds.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO200200927-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-EP07536.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR MPI; 2002-130908/17.  
 XX  
 PT Novel nucleic acid useful for diagnosis and therapy of diseases  
 PT associated with development genes such as diabetes, comprises a  
 PT sequence of a segment of chemically pretreated DNA of genes associated  
 PT with development  
 XX  
 PS Claim 1; SEQ ID NO 36; 27pp; English.  
 XX  
 CC The invention relates to a nucleic acid (I) comprising a sequence at  
 CC least 18 bases in length of a segment of chemically pretreated DNA (II)  
 CC of genes associated with development selected from 87 genes listed in

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:03:20 ; Search time 20.9513 Seconds  
(without alignments)  
2257.229 Million cell updates/sec

Title: US-09-477-082-31  
Perfect score: 21  
Sequence: 1 tagggatttgagattgtga 21

Scoring table: IDENTITY-NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N.Geneseq\_101002:\*

1:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	AAA51820	Sense primer for C
2	18.4	87.6	5518	AB154318	Chemically treated
3	18.4	87.6	5518	ABN80019	Human chemically m
4	18.4	87.6	5518	ABJ32217	Human immune syste
5	17.8	84.8	21	AAA51818	Sense primer for C
6	17.8	84.8	2204	AB122738	Drosophila melanog
7	17.8	84.8	2361	AB112786	Drosophila melanog
8	17.8	84.8	3536	AB113136	Drosophila melanog
9	17.8	84.8	13377	AAA54475	Tumour suppressor

10	17.8	84.8	13377	24	ABJ33462	Human immune syste
11	17	81.0	16914	24	AB170316	Chemically treated
12	17	81.0	16914	24	AA61254	Human gene regulat
13	16.8	80.0	898	21	AA652475	Arabidopsis thalia
14	16.8	80.0	1199	23	AB111065	Drosophila melanog
15	16.8	80.0	2325	21	AAA57385	Pea p12 gene 11th
16	16.8	80.0	3343	23	AB111064	Drosophila melanog
17	16.8	80.0	3441	21	AAA97382	Pea 11gnt-repressi
18	16.8	80.0	3916	22	AA68032	Human immune/haema
19	16.8	80.0	6127	24	AB132592	Human immune syste
20	16.8	80.0	168575	22	AAH21613	Human hypoxerlin r
21	16.4	78.1	519	24	ABQ36984	Oligonucleotide fo
22	16.4	78.1	519	24	ABQ36985	Oligonucleotide fo
23	16.4	78.1	588	23	ABV55682	Human prostate exp
24	16.4	78.1	981	23	AA553751	Helicobacter pylori
25	16.4	78.1	11358	22	AAH22434	P450RAI-2 containi
26	16.2	77.1	422	22	AA692173	Rabbit antiimmunogl
27	16.2	77.1	452	22	AA541743	Genomic sequence #
28	16.2	77.1	452	22	AAK11554	Human immune/haema
29	16.2	77.1	452	22	AAK80621	Human immune/haema
30	16.2	77.1	640	24	ABQ37686	Oligonucleotide fo
31	16.2	77.1	640	24	ABQ37687	Oligonucleotide fo
32	16.2	77.1	810	23	ABV10247	Human prostate exp
33	16.2	77.1	1587	21	AA642056	Arabidopsis thalia
34	16.2	77.1	3227	22	AAK44254	Human immune/haema
35	16.2	77.1	5310	24	ABK33934	Human DNA for stag
36	16.2	77.1	6408	24	AB133120	Human immune syste
37	16.2	77.1	7763	24	AB134437	Human immune syste
38	16.2	77.1	9927	24	AB132113	Human immune syste
39	16.2	77.1	11046	24	ABK31536	Signal transductio
40	16.2	77.1	11996	24	AB134493	Human metastasis a
41	16.2	77.1	15765	22	ABAI7911	Human nervous syst
42	16.2	77.1	15765	22	ABAI8250	Human nervous syst
43	16.2	77.1	15765	22	ABAI8276	Human nervous syst
44	16.2	77.1	15765	22	ABAI9009	Human nervous syst
45	16.2	77.1	16772	22	ABA20862	Human nervous syst

## ALIGNMENTS

RESULT 1						
ID	AAA51820	standard:	DNA:	21	BP.	
AC	AAA51820;					
XX						
AC	AAA51820;					
XX						
DT	31-OCT-2000	(first entry)				
DE	Sense primer for CASP8 unmethylated-specific disulfite treated DNA.					
XX						
XX	CASP8; caspase-8; promoter; inactivation; methylation; cysteine protease;					
KW	tumour suppressor; Chromosome 2q33-34; neuroblastoma; cancer;					
KW	death receptor; apoptosis; cytostatic; gene therapy; primer; ss.					
OS	Homo sapiens.					
XX						
PN	WO200039347-A1.					
XX						
PD	06-JUL-2000.					
XX						
PF	30-DEC-1999;	99WO-US31280.				
XX						
PR	31-DEC-1998;	98US-0114308.				
XX						
PA	(SUID-) ST JUDE CHILDREN'S RES HOSPITAL.					
XX						
PI	Kidd VJ, Lahti JM, Teltz T;					
XX						
DR	WPI; 2000-452423/39.					
XX						
PT	Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or					
PT	prognosing cancer, comprises detecting a modification of genomic DNA					

```

repeat_region 19590..20045
/note="LIM4 repeat: matches 2859..3326 of consensus"
repeat_region 20099..20355
/note="LIM4 repeat: matches 3603..3874 of consensus"
misc_feature 20275..20787
/note="match: GSS: Em:AQ547282"
repeat_region 21667..22029
/note="LIM1B repeat: matches 4..375 of consensus"
repeat_region 22560..22646
/note="L2 repeat: matches 2624..2710 of consensus"
repeat_region 23310..23954
/note="L2 repeat: matches 1993..2643 of consensus"
misc_feature complement(23742..24153)
/note="match: GSS: Em:AQ182243"
misc_feature 24153..24603
/note="match: GSS: Em:AQ880736"
misc_feature 24173..24662
/note="match: GSS: Em:AQ49802"
repeat_region 26631..26684
/note="27 copies 2 mer aa 74% conserved"
repeat_region 26986..27049
/note="32 copies 2 mer tt 67% conserved"
repeat_region 27058..28403
/note="LIP1A16 repeat: matches 4771..6156 of consensus"
repeat_region 28404..28699
/note="LIM4 repeat: matches 6..304 of consensus"
repeat_region 28700..29162
/note="LIP1A16 repeat: matches 4315..4771 of consensus"
repeat_region 29948..30441
/note="LIM1D repeat: matches 4..505 of consensus"
misc_feature complement(30606..31171)
/note="match: GSS: Em:AQ280648"
repeat_region 30764..30992
/note="L1 repeat: matches 4435..4668 of consensus"
repeat_region 31560..33433
/note="LIM9 repeat: matches 4436..6306 of consensus"
repeat_region 33796..34254
/note="L2 repeat: matches 2073..2570 of consensus"
repeat_region 34418..34526
/note="MIR repeat: matches 31..139 of consensus"
repeat_region 34653..34951
/note="LIM1 repeat: matches 5827..6105 of consensus"
repeat_region 34952..35443
/note="LOR1 repeat: matches 1..497 of consensus"
repeat_region 35444..35913
/note="LIM1 repeat: matches 5285..5827 of consensus"
repeat_region 35955..36631
/note="LIM1 repeat: matches 4545..5298 of consensus"
repeat_region 36633..37090
/note="LIM4 repeat: matches 3548..4026 of consensus"
repeat_region 37269..37376
/note="LIM4 repeat: matches 3131..3242 of consensus"
repeat_region 37510..37705
/note="LIMC repeat: matches 1813..2005 of consensus"
repeat_region 37760..38650
/note="LIM6 repeat: matches 5231..6088 of consensus"
repeat_region 38649..38823
/note="L1 repeat: matches 4553..4718 of consensus"
repeat_region 38824..39051
/note="PTR3 repeat: matches 743..965 of consensus"
repeat_region 39131..39181
/note="LIM30 repeat: matches 1..52 of consensus"
repeat_region 39185..39292
/note="PTR5 repeat: matches 1570..1679 of consensus"
repeat_region 39293..39981
/note="LIM12 repeat: matches 1..671 of consensus"
repeat_region 39982..40184
/note="L1 repeat: matches 4715..4912 of consensus"
repeat_region 40338..40650
/note="LIMX repeat: matches 1..306 of consensus"
repeat_region 40866..41242
/note="LIM16 repeat: matches 61..445 of consensus"
repeat_region 41259..41576

```

```

repeat_region /note="L2 repeat: matches 1777..2128 of consensus"
41824..41926
/note="LIP1A16 repeat: matches 6041..6143 of consensus"
repeat_region 42069..42189
/note="LIM1C repeat: matches 7..126 of consensus"
misc_feature 43350..43736
/note="match: GSS: Em:B60315"
misc_feature 43373..43929
/note="match: GSS: Em:AQ269706"
misc_feature 44214..44643
/note="match: GSS: Em:AQ358471"
misc_feature 44232..44969

```

```

Query Match 84.8%; Score 17.8; DB 9; Length 98368;
Best Local Similarity 90.5%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 TAGGGATTGGATTGCA 21
DB 50285 TAGGGATTGGATTGCA 50305

```

```

Search completed: July 8, 2003, 04:02:46
Job time : 79.7495 secs

```

19253..19416,19674..19873,19922..19960))  
 /gene="Y39D8B.3"  
 /ecodon\_start=1  
 /product="Hypothetical protein Y39D8B.3"  
 /protein\_id="AA060533.1"  
 /db\_xref="GI:7331845"  
 /translation="MTAYEIKKHNDITLIRBEDVADVDDDDMERKYNITLTH  
 GNTLITISDEARDLTKHMDQSLAGLGAATVAVINRKSKSEKEHTYCIASK  
 NYTAHAKCVTVLDQYKRNKLEKSTASLKHQQRSDGALKNSMNEYEY  
 KSGGEFVYKSGSGFVDEMEKROAVIREVQTRKSYELDNKSLPLAIARLT  
 ELYRAGNKEKPEKRMQVIOEIKNSTRIKGSONKMKRFGKFIETMGGLNP  
 DKAMOSIGMEDLFADPELITSEKEDAKDKEMSPDKILNPEITLRAIKIGMAA  
 GTRKDEANSDKKIALISQFQPSITLPEVANNITSLISITSLIGESDDPROSLT  
 KAKLLESTGEEMNFEVJASGVETVANKRKAREEERKMDQVDRKGPFTS  
 KENATOIGVEYAGKLELLDGFYKSLSSQGRTPKWKFLGYHRNSRPFQNFSEKIF  
 GISRKYORCHKLFRSPKFEPTPKFFPKFFNPKFQMESKNKGYTVMNDQLETTIGH  
 GSPFNSEALQKFRGISPSMPARIEIRMTAFREMKFDISRYKYE"

BASE COUNT 6587 a 3794 c 3934 g 5988 t  
 Query Match 84.8%; Score 17.8; DB 3; Length 20303;  
 Best Local Similarity 90.5%; Pred. No. 2e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGATTGGAGATTGCA 21  
 ||||||||||||| ||| ||  
 Db 6210 TAGGGATTGGAGATTGAGA 6230

RESULT 15  
 AL136453 98368 bp DNA linear PRI 31-MAR-2000  
 LOCUS AL136453  
 DEFINITION Human DNA sequence from clone RP1-249K20 on chromosome 10. Contains  
 STS and GSSs, complete sequence.  
 ACCESSION AL136453  
 VERSION AL136453.3 GI:6912131  
 KEYWORDS HTG.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 98368)  
 Graffham D.  
 Direct Submission  
 Submitted (14-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 Requests: clonerequest@sanger.ac.uk  
 On Feb 7, 2000 this sequence version replaced gi:6723337.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence has been finished according to sequence map criteria  
 as follows. An attempt is made to resolve all sequencing problems,  
 such as compressions and repeats, but not necessarily within known  
 annotated human repeat sequence elements (e.g. Alu), where the  
 sequence is ambiguous, there is an annotation using the 'unsure'  
 feature key.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information  
 on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/Celegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 10, constructed by the Sanger Centre Chromosome 10  
 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Ch10  
 RP1-249K20 is from the library RPCT-1 constructed at the Roswell  
 Park Cancer Institute by the group of Pieter de Jong. For further  
 details see http://bacpac.med.buffalo.edu/  
 VECTOR: pCYPAC2

# FEATURES

source

This sequence is the entire insert of clone RP1-249K20 The true  
 left end of clone RP4-53118 is at 61968 in this sequence.

1..98368  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP1-249K20"  
 /clone\_lib="RC1-1"  
 /complement(136..800)  
 /note="match: GSS: Em:A0779626"  
 154..858  
 /note="match: GSS: Em:A0009393"  
 164..868  
 /note="match: GSS: Em:A0009392"  
 208..592  
 /note="match: GSS: Em:B86219"  
 /complement(364..800)  
 /note="match: GSS: Em:A0465238"  
 2971..2996  
 /note="13 copies 2 mer aa 100% conserved"  
 3274..3309  
 /note="9 copies 4 mer caa 97% conserved"  
 4106..4317  
 /note="MIR repeat: matches 15..226 of consensus"  
 7363..7540  
 /note="LIM1 repeat: matches 5381..5563 of consensus"  
 /complement(7577..7983)  
 /note="match: GSS: Em:A0755541"  
 7606..7798  
 /note="MIR repeat: matches 2..192 of consensus"  
 7880..7956  
 /note="L2 repeat: matches 2666..2747 of consensus"  
 9022..9252  
 /note="AluX repeat: matches 1..231 of consensus"  
 9089..9457  
 /note="match: STS: Em:LL7677"  
 9279..9400  
 /note="61 copies 2 mer aa 63% conserved"  
 9284..9371  
 /note="8 copies 11 mer 68% conserved"  
 10062..10217  
 /note="MIR repeat: matches 57..221 of consensus"  
 11677..11813  
 /note="Alu repeat: matches 165..301 of consensus"  
 12586..12586  
 /note="LIM4 repeat: matches 4856..5160 of consensus"  
 12899..13407  
 /note="LIM4 repeat: matches 4886..5403 of consensus"  
 13271..13789  
 /note="match: GSS: Em:A0270786"  
 13417..13565  
 /note="LIM4 repeat: matches 6043..6289 of consensus"  
 13736..14644  
 /note="LIM4 repeat: matches 5417..6288 of consensus"  
 14365..14712  
 /note="match: STS: Em:Z53533"  
 14732..15046  
 /note="MER34 repeat: matches 190..534 of consensus"  
 15047..16331  
 /note="LIM4 repeat: matches 514..1735 of consensus"  
 16356..16583  
 /note="LIM4 repeat: matches 3909..4135 of consensus"  
 17777..17838  
 /note="MER63 repeat: matches 979..1046 of consensus"  
 18114..18346  
 /note="MIR repeat: matches 11..249 of consensus"  
 /complement(18285..18763)  
 /note="match: STS: Em:G63247"  
 18450..19006  
 /note="match: GSS: Em:A0387926"  
 19193..19396  
 /note="LIMC repeat: matches 952..1353 of consensus"

VERSION ACO06749.1 GI:4263189  
 KEYWORDS HTG.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 20303)  
 AUTHORS Waterston, R.  
 TITLE Genome sequence of the nematode *C. elegans*: a platform for  
 investigating biology. The *C. elegans* Sequencing Consortium  
 JOURNAL Science 282 (5396), 2012-2018 (1998)  
 MEDLINE 98069613  
 PUBMED 9851916  
 REFERENCE 2 (bases 1 to 20303)  
 AUTHORS Becker, M., Graves, T., and Yeakum, M.  
 TITLE The sequence of *C. elegans* cosmid Y39DB8  
 JOURNAL Unpublished (2001)  
 REFERENCE 3 (bases 1 to 20303)  
 AUTHORS Waterston, R. H.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-FEB-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE 4 (bases 1 to 20303)  
 AUTHORS Waterston, R. H.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-MAR-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE 5 (bases 1 to 20303)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-MAR-2000) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 REFERENCE 6 (bases 1 to 20303)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-MAY-2002) Department of Genetics, Washington  
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
 Louis, MO 63110, USA  
 COMMENT Submitted by:  
 Genome Sequencing Center  
 Department of Genetics, Washington University  
 St. Louis, MO 63110, USA, and  
 Sanger Centre, Hinxton Hall  
 Cambridge CB10 1RQ, England  
 email: r.w@nematoe.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.  
 It may be shorter because we only sequence overlapping sections  
 once, or longer because we provide a small overlap between  
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all  
 regions were double stranded, sequenced with an alternate chemistry  
 or covered by high quality data (i.e., phred quality >= 30); an  
 attempt was made to resolve all sequencing problems, such as  
 compressions and repeats; all regions were covered by sequence from  
 more than one m13 subclone.

For a graphical representation of this cosmid sequence and its  
 analysis see:  
<http://www.wormbase.org/db/seq/sequence?name=Y39DB8;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is T22H9, 3951 bp overlap; the 3' cosmid is 2K6, 200  
 bp overlap. Actual start of this cosmid is at base position 31064  
 of Y39DB8; actual end is at 27075 of F48G7.

NOTES:

Coding sequences below are the result of integration and manual

review of the following data: computer analysis using the program  
 GeneFinder (P. Green and L. Hillier, personal communication), the  
 large scale EST projects of Yui Kohara  
 ([http://www.ddbj.nig.ac.jp/c-elegans/html/CE\\_INDEX.html](http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html)) and The *C.*  
*elegans* ORFome cloning project (<http://wofdb.dcfi.harvard.edu/>),  
 similarity to other proteins from Blastx analyses  
 (<http://blast.wustl.edu/>), sequence conservation with *C. briggsae*  
 using Jim Kent's WABA alignment program (Genome Research  
 10:115-1125, 2000), individual *C. elegans* GenBank submissions,  
 and personal communications with *C. elegans* researchers. TRANS  
 are predicted using the program TRANScan-SE (Lowe, T.M. and  
 Eddy, S.R., 1997, Nucleic Acids. Res., 25, 955-964).

Location/Qualifiers

1. 20303  
 /organism="Caenorhabditis elegans"  
 /strain="Bristol N2"  
 /db\_xref="taxon:6239"  
 /chromosome="v"  
 /clone="Y39DB8"  
 5238..8372  
 /gene="Y39DB8.2"  
 /note="for a graphical representation of this gene see:  
<http://www.wormbase.org/db/seq/sequence?name=Y39DB8.2;class=Sequence>"  
 join(5238..5295,5343..5596,5649..5845,6631..6794,  
 6844..7138,7183..7333,7462..7716,7763..8101,8157..8372)  
 /gene="Y39DB8.2"  
 /codon\_start=1  
 /product="Hypothetical protein Y39DB8.2"  
 /protein\_id="AA060532.1"  
 /db\_xref="GI:7331844"  
 /translation="MGFTFRFLFVYFVGVRSVSGSIISREVPVSTPPQOSTLL  
 PFWLWENKPEFTSKPKRNALREGLIOVRGKIPKELAPVNTKIDENKKE  
 ODTLEITRDEPDADVDDVDEHVRVNTLTGDKNTITITSDPARDDLYKRWMSGLA  
 GIMGAVTVIVIKRKMTSEKDRHICINESKNVTAHAKCYVYVLDQNNKNEKLYK  
 STLATRVSRRKTRADNDITKTRLDHDKTVSGASEYRVRKAGESIPDEMKRQ  
 AAIRSNVQTRKRYTIRKNSLSPALILARKLELVRAKNGKQEPKQVQVIOEIK  
 SFRIGKNNKRRKRSKFSVSTKKEAKLNGMEDLEDDPILISEMIL  
 KKRKMEERMSDDKIMEPIKILIREAKIGVAMGKSGEIGDKVALLSPQMSIL  
 PDEVANDTSLSPSILSHGQGTDESLTKRLPLMEDGQEWNNFVLEASGV  
 ETVEKRIAEKEDEDERKDFVCDGSRPFTKENVETIYEATKIDILINGYKT  
 MSKDOLDALNSGYSIMDDKQIDMLYGGSPYNNSEALDKRGIPLEMPRIQDIR  
 LLAEDMKTELSKRVRLI"  
 10676..13543  
 /gene="Y39DB8.1"  
 /note="for a graphical representation of this gene see:  
<http://www.wormbase.org/db/seq/sequence?name=Y39DB8.1;class=Sequence>"  
 join(10676..10709,11442..11472,11529..12042,12415..12517,  
 12562..12680,12730..12900,12963..13130,13280..13543)  
 /gene="Y39DB8.1"  
 /codon\_start=1  
 /product="Hypothetical protein Y39DB8.1"  
 /protein\_id="AA060531.1"  
 /db\_xref="GI:7331843"  
 /translation="MNFTEGAVYGLDNTSSSDIQSVRLKHQLTGTPSKLPLFLN  
 LIAVAKERSHNSGLQFIVDDYKRGKIDKDPVLAINMTDGNKTVTISEEAYOL  
 LKAVDOASGCPMALMSITKDKRLDLEPRHILQHNKNQVAKSIFEAQCVVTLIKRL  
 DOQRIRIKRKLKFRSOLRRARAKRDLPEYRDKRVVQKQOYLKKKEKSTP  
 FSLIKRLTIDYRKNKKEPKKKNQVYINIDELNRIOKKKKQDDQDFPNVSKYS  
 RPKMKLGINPERRARAGDSIGETISANDTIVLSPFALHSSSELDQMTS  
 LTGVNSVNDKIKALMSQFMSVLPDEQANDTVLLSPFALHSSSELDQMTS  
 RAMPRLSENGHDDMNFLVLEAGVLEAKMFEDRKQEMEFRRHESSENGQVET  
 RONVSEMYGDYTSKIDSMEKLRKMSAAMHNNKSTGYALNSOQISEFPQSPYN  
 DSHAVENYENLRNRNDIPDLENNIHQAKREDAFVAROVGHELMRTTHYKNFCISRK  
 M"  
 Complement(15523..19960)  
 /gene="Y39DB8.3"  
 /note="for a graphical representation of this gene see:  
<http://www.wormbase.org/db/seq/sequence?name=Y39DB8.3;class=Sequence>"  
 Complement(join(15523..15730,15925..16058,16137..16475,  
 17187..17311,17828..17954,18005..18155,18905..19199,

CDS

gene

CDS

gene

CDS

gene

FEATURES  
 source

```

misc_feature      /note="assembly-fragment"
                  120974..136795
misc_feature      /note="assembly-fragment"
                  136896..151116
misc_feature      /note="assembly-fragment"
                  151217..179686
BASE COUNT      58041 a 31435 c 31392 g 57013 t 1805 others
ORIGIN
Query Match      85.7%; Score 18; DB 2; Length 179686;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TAGGGATTGGAGATTG 18
        ||||||||||||||||
Db      6412 TAGGGATTGGAGATTG 6395

RESULT 11
LOCUS      AK096637      4369 bp      mRNA      linear      PRI 15-JUL-2002
DEFINITION Homo sapiens CDNA FLJ39318 fls, clone OCBP2014089.
ACCESSION  AK096637.1 GI:21756173
VERSION     AK096637.1 GI:21756173
KEYWORDS    cllgo capping; fls (full insert sequence);
SOURCE      Homo sapiens fetal brain CDNA to mRNA, clone_l1b:OCBPF2
            clone:OCBPF2014089.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS     1 Nishi,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
            Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Makamatsu,A., Ishii,S.,
            Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saico,K., Nishikawa,T.,
            Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
            Kikuchi,H., Kanda,K., Wagatsuma,M., Kurakawa,K., Kanehori,K.,
            Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
            Sugano,S., Nagahari,K., Masuhio,Y., Nagai,K. and Isogai,T.
            NEDO human CDNA sequencing project
TITLE       Unpublished
JOURNAL     2 (bases 1 to 4369)
REFERENCE   Isogai,T. and Yamamoto,J.
AUTHORS     Direct Submission
TITLE       Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
JOURNAL     Kazusa-Kamatari, Kisarazu, Chiba 292-0912, Japan
            (E-mail:genomise@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human CDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; CDNA full insert sequencing:
            Research Association for Biotechnology (RAB); CDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing; RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES
source
1. .4369
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="OCBPF2014089"
/tissue_type="brain"
/clone_l1b="OCBPF2"
/dev_stage="fetal"
/note="cloning vector: PME18SFL3"
BASE COUNT      1163 a 920 c 1044 g 1242 t
ORIGIN
Query Match      84.8%; Score 17.8; DB 9; Length 4369;
Best Local Similarity 90.5%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TAGGGATTGGAGATTGTA 21
        |||||
Db      1 TAGGGATTGGAGATTGTA 21

```

```

Db      1369 TAGGGCTTTGGAGATTGTA 1389

RESULT 12
LOCUS      AX251229      13377 bp      DNA      linear      PAT 05-OCT-2001
DEFINITION Sequence 197 from Patent WO018912.
ACCESSION  AX251229
VERSION     AX251229.1 GI:15984652
KEYWORDS
SOURCE      synthetic construct.
ORGANISM    synthetic construct
            artificial sequences.
REFERENCE   1 (bases 1 to 13377)
AUTHORS     Olek,A., Piepenbrock,C. and Berlin,K.
TITLE       Diagnosis of diseases associated with tumor suppressor genes and
            oncogenes
JOURNAL     Patent: WO 0168912-A 197 20-SEP-2001;
            Epigenomics AG (DE)
FEATURES
source
1. .13377
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT      3915 a 146 c 3046 g 6270 t
ORIGIN
Query Match      84.8%; Score 17.8; DB 6; Length 13377;
Best Local Similarity 90.5%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TAGGGATTGGAGATTGTA 21
        ||||||||||||||||
Db      1704 TAGGGATTGGAGATTGTA 1724

RESULT 13
LOCUS      AX346364      13377 bp      DNA      linear      PAT 01-FEB-2002
DEFINITION Sequence 1435 from Patent WO0200928.
ACCESSION  AX346364
VERSION     AX346364.1 GI:18494250
KEYWORDS
SOURCE      synthetic construct.
            synthetic construct
            artificial sequences.
ORGANISM    Olek,A., Piepenbrock,C. and Berlin,K.
TITLE       Diagnosis of diseases associated with the immune system
JOURNAL     Patent: WO 0200928-A 1435 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES
source
1. .13377
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT      3915 a 146 c 3046 g 6270 t
ORIGIN
Query Match      84.8%; Score 17.8; DB 6; Length 13377;
Best Local Similarity 90.5%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TAGGGATTGGAGATTGTA 21
        ||||||||||||||||
Db      1704 TAGGGATTGGAGATTGTA 1724

RESULT 14
LOCUS      AC006749      20303 bp      DNA      linear      INV 24-MAY-2002
DEFINITION Caenorhabditis elegans cosmid Y39DB8, complete sequence.
ACCESSION  AC006749

```





```

*      81943      82042: gap of unknown length
*      82043      150022: contig of 67980 bp in length
*      150023      150122: gap of unknown length
*      150123      152401: contig of 2279 bp in length
*      152402      152501: gap of unknown length
*      152502      158391: contig of 5890 bp in length
*      158392      158491: gap of unknown length
*      158492      152895: contig of 4404 bp in length
*      162896      162995: gap of unknown length
*      162996      168149: contig of 5154 bp in length
*      168150      168249: gap of unknown length
*      168250      176144: contig of 7895 bp in length
*      176145      176244: gap of unknown length
*      176245      186417: contig of 10173 bp in length
*      186418      186517: gap of unknown length
*      186518      199541: contig of 13024 bp in length
*      199542      199642: gap of unknown length
*      199642      214787: contig of 15146 bp in length.
FEATURES
source
1..214787
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-326J1"
1..15958
/misc_feature
/feature="assembly_name:Contig10"
16059..47159
/misc_feature
/feature="assembly_name:Contig11"
47260..81942
/misc_feature
/feature="assembly_name:Contig12"
82043..150022
/misc_feature
/feature="assembly_name:Contig13"
150123..152401
/misc_feature
/feature="assembly_name:Contig14"
152502..158391
/misc_feature
/feature="assembly_name:Contig15"
158492..162895
/misc_feature
/feature="assembly_name:Contig16"
162996..168149
/misc_feature
/feature="assembly_name:Contig17"
168250..176144
/misc_feature
/feature="assembly_name:Contig18"
176245..186417
/misc_feature
/feature="assembly_name:Contig19"
186518..199541
/misc_feature
/feature="assembly_name:Contig20"
199642..214787
misc_count 57078 a 49921 c 49241 g 57431 t 1116 others
ORIGIN
Query Match      87.6%; Score 18.4; DB 2; Length 214787;
Best Local Similarity 95.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
2 AGGGGATTGGAGATTGGA 21
Db      109370 AGGGGACTTGGAGATTGGA 109389

```

```

RESULT 9
LOCUS      AC091938      70122 bp      DNA      linear      PRI 31-MAY-2002
DEFINITION Homo sapiens chromosome 5 clone RP11-329N11, complete sequence.
ACCESSION      AC091938
VERSION      AC091938.2      GI:21281476
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 70122)
DOE Joint Genome Institute and Stanford Human Genome Center.

```

```

TITLE      Direct Submission
JOURNAL      Unpublished
REFERENCE
2 (bases 1 to 70122)
AUTHORS      DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL      Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 70122)
REFERENCE
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE      Direct Submission
JOURNAL      Submitted (31-MAY-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On May 31, 2002 this sequence version replaced gi:14333874.
COMMENT
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.5% of Sequence.
Estimated Total Number of Errors is 0.2.
NOTE: This insert is not the entire sequence of the clone (entire
sequence is 153kb). It is clipped at the overlaps with AC022139 and
AC109455. The number of bases overlapped with AC022139 is 12649 and
with AC109455 is 3639.
FEATURES
source
1..70122
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-329N11"
Location/Qualifiers
1..15958
/misc_feature
/feature="assembly_name:Contig10"
16059..47159
/misc_feature
/feature="assembly_name:Contig11"
47260..81942
/misc_feature
/feature="assembly_name:Contig12"
82043..150022
/misc_feature
/feature="assembly_name:Contig13"
150123..152401
/misc_feature
/feature="assembly_name:Contig14"
152502..158391
/misc_feature
/feature="assembly_name:Contig15"
158492..162895
/misc_feature
/feature="assembly_name:Contig16"
162996..168149
/misc_feature
/feature="assembly_name:Contig17"
168250..176144
/misc_feature
/feature="assembly_name:Contig18"
176245..186417
/misc_feature
/feature="assembly_name:Contig19"
186518..199541
/misc_feature
/feature="assembly_name:Contig20"
199642..214787
misc_count 23470 a 11786 c 12143 g 22723 t
ORIGIN
Query Match      85.7%; Score 18; DB 9; Length 70122;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 TAGGGGATTGGAGATTG 18
Db      27627 TAGGGGATTGGAGATTG 27644

```

```

RESULT 10
LOCUS      AC023987/c      179686 bp      DNA      linear      HTG 24-AUG-2002
DEFINITION Homo sapiens chromosome 5 clone RP11-390G24 map 5, WORKING DRAFT
SEQUENCE, 19 unordered pieces.
ACCESSION      AC023987
VERSION      AC023987.3      GI:7705153
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 179686)
AUTHORS      Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE      Homo sapiens chromosome 5, clone RP11-390G24
JOURNAL      Unpublished
2 (bases 1 to 179686)
AUTHORS      Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bedalov, F., Boguslavsky, L.,
Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,
Fenstermaker, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,
Galagan, J., Gardy, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howard, J., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Landers, T., Laroque, K., Lechoczy, J., Levine, R.,
Lieu, C., Liu, G., Locke, K., MacDonald, P., Marguis, N., McCarthy, M.,
McKean, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J.,
Menkus, L., Mironov, T., Miranda, C., Munga, V., Morrow, J., Naylor, J.,
Norman, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M.,
Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C.,

```

```

repeat_region complement(4063, .4146)
/rpt_family="MIR"
repeat_region complement(4701, .4881)
/rpt_family="MIR"
repeat_region 4959, .4986
/rpt_family="AT-rich"
repeat_region 5136, .5373
/rpt_family="MTL1"
repeat_region complement(5374, .5639)
/rpt_family="Alusx"
repeat_region 5640, .5876
/rpt_family="MTL1"
repeat_region complement(6582, .6683)
/rpt_family="MIR"
repeat_region 7263, .7329
/rpt_family="Charlie8"
repeat_region complement(7333, .7638)
/rpt_family="Alusq"
repeat_region 7723, .7847
/rpt_family="Charlie8"
repeat_region 7940, .8083
/rpt_family="Charlie8"
repeat_region complement(8330, .8639)
/rpt_family="Alusx"
repeat_region complement(8649, .8683)
note="Single clone coverage"
8844, .9152
/rpt_family="Aluy"
repeat_region 9242, .9401
/rpt_family="L1PBI"
repeat_region 9402, .9586
/rpt_family="Aluy"
repeat_region 9587, .9664
/rpt_family="Alusq/q"
repeat_region complement(9914, .10340)
/rpt_family="L2"
repeat_region 10749, .10878
/rpt_family="Charlie8"
repeat_region complement(11241, .11322)
/rpt_family="L2"
repeat_region 11651, .11859
/rpt_family="Alusx"
repeat_region 12129, .12154
/rpt_family="AT-rich"
repeat_region complement(12321, .12638)
/rpt_family="Alusq"
repeat_region complement(13500, .13642)
/rpt_family="MIR3"
repeat_region complement(13659, .13955)
/rpt_family="Alusq"
repeat_region complement(13973, .14204)
/rpt_family="MIR"
repeat_region 14345, .14369
/rpt_family="(CAAA)n"
repeat_region 14693, .14753
/rpt_family="(TTTA)n"
repeat_region complement(14758, .15037)
/rpt_family="Alusx"
repeat_region 17097, .17142
/rpt_family="(CA)n"
repeat_region 17949, .18097
/rpt_family="L1ME4A"
repeat_region complement(18645, .18755)
/rpt_family="MIR"
repeat_region 18982, .19069
/rpt_family="MIR"
repeat_region 19071, .19211
/rpt_family="F1AM_C"
repeat_region 19214, .19339
/rpt_family="MIR"
repeat_region complement(19411, .19606)
/rpt_family="MIR"
repeat_region 19725, .19777

```

```

repeat_region /rpt_family="MIR"
20135, .20247
/rpt_family="MIR3"
repeat_region complement(20248, .20536)
/rpt_family="Alusx"
repeat_region 20537, .20556

Query Match 87.6% Score 18.4; DB 9; Length 188868;
Best Local Similarity 95.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 AGGGGATTGGAGATTGTGA 21
|||||
Db 2388 AGGGGATTGAAGATTGTGA 2369

RESULT 8
AC123853 214787 bp DNA linear HTG 01-JUN-2002
LOCUS Mus musculus chromosome UNK clone RP23-326J1, WORKING DRAFT
DEFINITION
ACCESSION AC123853
VERSION AC123853.1 GI:21307546
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 214787)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 214787)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watscn.wustl.edu
Project Information
Center project name: M_BA0326J01
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 207948 bases at least Q40
Consensus quality: 208444 bases at least Q30
Consensus quality: 208788 bases at least Q20
Insert size: 171000; agarose-ff
Insert size: 213687; sum-of-contigs
Quality coverage: 11.36 in Q20 bases; agarose-ff
Quality coverage: 9.45 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 15958: contig of 15958 bp in length
* 15959 16058: gap of unknown length
* 16059 47159: contig of 31101 bp in length
* 47160 47259: gap of unknown length
* 47260 81942: contig of 34683 bp in length

```

```

misc.feature      /note="assembly_name:Contig3"
                  136514..140443
misc.feature      /note="assembly_name:Contig4"
                  140544..148908
misc.feature      /note="assembly_name:Contig6"
                  149009..160432
misc.feature      /note="assembly_name:Contig7"
                  160533..171996
misc.feature      /note="assembly_name:Contig8"
                  172097..186048
misc.feature      /note="assembly_name:Contig9"
                  186049..196048
BASE COUNT      50542 a 44621 c 41477 g 48405 t 1003 others
ORIGIN
Query Match      87.6% Score 18.4; DB 2; Length 186048;
Best Local Similarity 95.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY              2 AGGCGATTGGAGATTTGTA 21
Db              72797 AGGCGATTGGAGATTTGTA 72778

```

```

RESULT 7
AC023232/c      188868 bp      DNA      linear      PRI 01-JUL-2002
LOCUS          Homo sapiens chromosome 11, clone RP11-793111, complete sequence.
AC023232
AC023232.16      GI:21637520
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE
AUTHORS
1 (bases 1 to 188868)
Birken,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 11, clone RP11-793111
Unpublished
2 (bases 1 to 188868)
Birken,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Farescor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heatord,A., Horton,L.,
Howard,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Mackdonald,P., Margus,N., McEwan,P., McGurk,A., McKernan,K.,
McHeaters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Piere,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.T.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 188868)
Birken,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouckgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Margus,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
Meneus,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C.,

```

```

TITLE
JOURNAL
COMMENT
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T., Peterson,K., Phunkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testfaye,S.,
Theodore,J., Topham,K., Travers,M., Travis,N., Trigglio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (26-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 188868)
Birken,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouckgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenka,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Testfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 1, 2002 this sequence version replaced gi:21591935.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
-----Project Information-----
Center project name: L6735
Center clone name: 793_111
Location/Qualifiers
1..188868
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone_id="RP11-793111"
/clone_lib="RPC1-11 Human Male BAC"
585..639
/rpt_family="(A)n"
1098..1295
/rpt_family="MIR"
1365..1584
/rpt_family="MIR"
2471..2546
/rpt_family="L2"
2700..2848
/rpt_family="L2"
2860..3079
/rpt_family="L2"
3214..3347
/rpt_family="FLAM_C"
complement(3080..3213)
/rpt_family="FLAM_C"
complement(3363..3677)
/rpt_family="Alusp"

```

## FEATURES

```

source
1..188868
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone_id="RP11-793111"
/clone_lib="RPC1-11 Human Male BAC"
585..639
/rpt_family="(A)n"
1098..1295
/rpt_family="MIR"
1365..1584
/rpt_family="MIR"
2471..2546
/rpt_family="L2"
2700..2848
/rpt_family="L2"
2860..3079
/rpt_family="L2"
3214..3347
/rpt_family="FLAM_C"
complement(3080..3213)
/rpt_family="FLAM_C"
complement(3363..3677)
/rpt_family="Alusp"

```

```

RESULT 5
AC121986/c 182686 bp DNA linear ROD 17-AUG-2002
LOCUS AC121986
DEFINITION Mus musculus chromosome 10 clone RP24-29317, complete sequence.
AC121986
VERSION AC121986.2 GI:22296788
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 182686)
JOURNAL McPherson,J.D. and Waterston,R.H.
REFERENCE The sequence of Mus musculus clone
AUTHORS unpublished
TITLE 2 (bases 1 to 182686)
JOURNAL McPherson,J.D. and Waterston,R.H.
REFERENCE Direct Submission
AUTHORS Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
TITLE Parkway, St. Louis, MO 63108, USA
JOURNAL 3 (bases 1 to 182686)
REFERENCE McPherson,J.D. and Waterston,R.H.
AUTHORS Direct Submission
TITLE Submitted (17-AUG-2002) Genome Sequencing Center, 4444 Forest Park
JOURNAL Parkway, St Louis, MO 63108, USA
COMMENT On Aug 17, 2002 this sequence version replaced gi:21040110.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BB0293107
-----
FEATURES
source
1. 182686
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="10"
/clone="RP24-29317"
BASE COUNT 52709 a 37532 c 36957 g 55468 t
ORIGIN
Query Match 87.6%; Score 18.4; DB 10; Length 182686;
Best Local Similarity 95.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0.

QY 2 AGGGGATTGGAGATTGCA 21
|||||
Db 86257 AGGGGATTGGAGATTGCA 86238

RESULT 6
LOCUS AC074190/c 186048 bp DNA linear HTG 10-SEP-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-465124, WORKING DRAFT
AC074190
ACCESSION AC074190
VERSION AC074190.3 GI:10048050
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 186048)
JOURNAL Waterston,R.H.
REFERENCE The sequence of Homo sapiens clone
AUTHORS unpublished
TITLE 2 (bases 1 to 186048)
JOURNAL Waterston,R.H.
REFERENCE Direct Submission

```

```

JOURNAL
Submitted (16-Jul-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 10, 2000 this sequence version replaced gi:9931959.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project information -----
Center project name: H.NH0465124
Summary Statistics
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ER; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 179088 bases at least Q40
Consensus quality: 181591 bases at least Q30
Consensus quality: 182797 bases at least Q20
Insert size: 206000; agarose-IP
Insert size: 185048; sum-of-contigs
Quality coverage: 5.59 in Q20 bases; agarose-IP
Quality coverage: 6.26 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
17943: contig of 17943 bp in length
17944 18043: gap of unknown length
18044 35118: contig of 17075 bp in length
35119 35219: gap of unknown length
35219 63426: contig of 28208 bp in length
63427 63526: gap of unknown length
63527 125447: contig of 61921 bp in length
125448 132197: gap of unknown length
132198 132297: contig of 6650 bp in length
132298 136413: gap of unknown length
136414 136513: contig of 4116 bp in length
136514 140443: gap of unknown length
140444 140543: contig of 3930 bp in length
140544 148908: gap of unknown length
148909 149008: contig of 8365 bp in length
149009 160432: gap of unknown length
160433 160532: contig of 11424 bp in length
160533 171996: gap of unknown length
171997 172096: contig of 11464 bp in length
172097 186048: gap of unknown length
186049 19352: contig of 13952 bp in length.
Location/Qualifiers
1. 186048
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="P11-465124"
1. 17943
/note="assembly_name:Contig10"
18044. 35118
/note="assembly_name:Contig11"
35219. 63426
/note="assembly_name:Contig12"
63527. 125447
/note="assembly_name:Contig13"
125548. 132197
/note="assembly_name:Contig5
clone_end:T7
vector_side:left"
132296. 136413

```

## JOURNAL Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 20, 2002 this sequence version replaced g1:17149713.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRB

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L21818

Center clone name: 2183\_M\_20

## FEATURES

source	Location/Qualifiers
	1. 136372
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="11"
	/map="11"
	/clone="CTD1-2183M20"
	/clone_lib="CTD1 Human BAC"
	/complement(597..704)
repeat_region	/rpt_family="MIR"
repeat_region	/rpt_family="MIR"
repeat_region	/rpt_family="MIR"
repeat_region	/rpt_family="MIR"
repeat_region	/rpt_family="AluSx"
repeat_region	4234..4365
repeat_region	/rpt_family="MIR3"
repeat_region	/complement(487..4953)
repeat_region	/rpt_family="MIR"
repeat_region	5349..5395
repeat_region	/rpt_family="MIR"
repeat_region	5420..5705
repeat_region	/rpt_family="AluSx"
repeat_region	5706..5726
repeat_region	/rpt_family="AT-rich"
repeat_region	/complement(6150..6222)
repeat_region	/rpt_family="Charlie1"
repeat_region	6267..6564
repeat_region	/rpt_family="AluSg"
repeat_region	/complement(6569..6974)
repeat_region	/rpt_family="Charlie1"
repeat_region	/complement(7026..7042)
repeat_region	/rpt_family="Charlie1"
repeat_region	7043..7422
repeat_region	/rpt_family="MER101"
repeat_region	7423..7732
repeat_region	/rpt_family="AluSx"
repeat_region	7733..7866
repeat_region	/rpt_family="MER101"
repeat_region	/complement(7867..7941)
repeat_region	/rpt_family="Charlie1"
repeat_region	/complement(7911..8074)
repeat_region	/rpt_family="Charlie1"
repeat_region	/complement(8122..8326)
repeat_region	/rpt_family="L2"
repeat_region	/complement(8450..8745)
repeat_region	/rpt_family="AluSg"
repeat_region	/complement(9009..9077)
repeat_region	/rpt_family="MIR"
repeat_region	9195..9269
repeat_region	/rpt_family="FLAM"
repeat_region	9276..9422
repeat_region	/rpt_family="AluCo"
repeat_region	9423..9659
repeat_region	/rpt_family="AluDo"
repeat_region	/complement(9819..9959)
repeat_region	/rpt_family="L2"

repeat_region	complement(10121..10231)
repeat_region	/rpt_family="MIR"
repeat_region	10740..10923
repeat_region	/rpt_family="MER5A"
repeat_region	/complement(10947..11191)
repeat_region	/rpt_family="L2"
repeat_region	11396..11619
repeat_region	/rpt_family="MIR"
repeat_region	11646..11775
repeat_region	/rpt_family="MIR3"
repeat_region	12444..12474
repeat_region	/rpt_family="(CA)n"
repeat_region	/complement(12619..12834)
repeat_region	/rpt_family="MER30"
repeat_region	/complement(12911..13257)
repeat_region	/rpt_family="MER58B"
repeat_region	13529..13839
repeat_region	/rpt_family="AluSx"
repeat_region	/complement(14058..14147)
repeat_region	/rpt_family="L2"
repeat_region	14721..15031
repeat_region	/rpt_family="AluSp"
repeat_region	/complement(15717..15794)
repeat_region	/rpt_family="MIR"
repeat_region	16155..16459
repeat_region	/rpt_family="AluCo"
repeat_region	16463..16762
repeat_region	/rpt_family="AluSg"
repeat_region	/complement(17316..17378)
repeat_region	/rpt_family="MIR"
repeat_region	17593..17900
repeat_region	/rpt_family="AluUb"
repeat_region	17934..17968
repeat_region	/rpt_family="AT-rich"
repeat_region	/complement(18398..18456)
repeat_region	/rpt_family="MIR"
repeat_region	19382..19682
repeat_region	/rpt_family="MSTP"
repeat_region	19683..19985
repeat_region	/rpt_family="AluSx"
repeat_region	19987..20282
repeat_region	/rpt_family="AluSc"
repeat_region	20283..20567
repeat_region	/rpt_family="AluY"
repeat_region	20568..20616
repeat_region	/rpt_family="MSTP"
repeat_region	/complement(21371..21672)
repeat_region	/rpt_family="AluSx"
repeat_region	22147..22445
repeat_region	/rpt_family="AluY"
repeat_region	22493..22790
repeat_region	/rpt_family="AluSx"
repeat_region	/complement(22969..23165)
repeat_region	/rpt_family="MIR"
repeat_region	23223..23264
repeat_region	/rpt_family="L2"
repeat_region	/complement(23265..23616)
repeat_region	/rpt_family="THE1B"
repeat_region	23617..23696
repeat_region	/rpt_family="L2"
repeat_region	25006..25336
repeat_region	/rpt_family="AluSx"
repeat_region	/complement(25722..25905)
repeat_region	/rpt_family="AluCo"

Query Match 87.6%; Score 18.4; DB 9; Length 136372;  
 Best Local Similarity 95.0%; Pred. No. 1.2e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGGGGATTGGAGATTGCA 21  
 DB 58131 AGGGGATTGAAGATTGCA 58112

## FEATURES

Location/Qualifiers  
1. 5518  
/organism="synthetic construct"

BASE COUNT 1383 a 142 c 1452 g 2541 t  
ORIGIN

Query Match 87.6%; Score 18.4; DB 6; Length 5518;  
Best Local Similarity 95.0%; Pred. No. 79;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TAGGGGATTGGAGATTGTG 20  
|||||  
46 TAGGGGATTGGAGATTGTG 65

RESULT 2  
AX344611 5518 bp DNA linear PAT 01-FEB-2002  
LOCUS  
DEFINITION Sequence 36 from Patent WO0200927.  
ACCESSION AX344611  
VERSION AX344611.1 GI:18492497

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1  
Olek, A., Piepenbrock, C. and Berlin, K.  
Diagnosis of diseases associated with development genes  
Patent: WO 0200927-A 36 03-JAN-2002;  
Epigenomics AG (DE)

FEATURES  
source  
1. 5518  
/organism="synthetic construct"

BASE COUNT 1383 a 142 c 1452 g 2541 t  
ORIGIN

Query Match 87.6%; Score 18.4; DB 6; Length 5518;  
Best Local Similarity 95.0%; Pred. No. 79;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TAGGGGATTGGAGATTGTG 20  
|||||  
46 TAGGGGATTGGAGATTGTG 65

RESULT 3  
AX345119 5518 bp DNA linear PAT 01-FEB-2002  
LOCUS  
DEFINITION Sequence 190 from Patent WO0200928.  
ACCESSION AX345119  
VERSION AX345119.1 GI:18493005

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1  
Olek, A., Piepenbrock, C. and Berlin, K.  
Diagnosis of diseases associated with the immune system  
Patent: WO 0200928-A 190 03-JAN-2002;  
Epigenomics AG (DE)

FEATURES  
source  
1. 5518  
/organism="synthetic construct"

BASE COUNT 1383 a 142 c 1452 g 2541 t  
ORIGIN

Query Match 87.6%; Score 18.4; DB 6; Length 5518;

Best Local Similarity 95.0%; Pred. No. 79;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TAGGGGATTGGAGATTGTG 20  
|||||  
46 TAGGGGATTGGAGATTGTG 65

RESULT 4  
AC103828/c 136372 bp DNA linear PRI 20-APR-2002  
LOCUS  
DEFINITION Homo sapiens chromosome 11, clone CTD1-2183M20, complete sequence.  
ACCESSION AC103828  
VERSION AC103828.2 GI:20219181

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1  
Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 136372)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Homo sapiens chromosome 11, clone CTD1-2183M20  
Unpublished  
2 (bases 1 to 136372)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,  
Anderson, S., Barina, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,  
Cooke, P., Develiano, K., Dewar, K., Diaz, J. S., Dodge, S., Fairo, S.,  
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagob, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K.,  
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
Maclean, C., MacDonald, P., Major, J., Margolis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPherson, R., Melchior, J.,  
Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Piere, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Riese, C., Rogov, P.,  
Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Sudramanian, A., Talamas, V., Testaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 136372)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,  
Anderson, S., Barina, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,  
Cook, A., Cooke, P., Develiano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
Fairo, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagob, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R.,  
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
Maclean, C., MacDonald, P., Major, J., Margolis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., Melchior, J., Meneus, L.,  
Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,  
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Piere, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Riese, C., Rogov, P.,  
Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Sudramanian, A., Talamas, V., Testaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

Direct Submission

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:07:09 ; Search time 74.5828 seconds

(without alignments)  
8194.363 Million cell updates/sec

Title: US-09-477-082-31

Perfect score: 21

Sequence: 1 tagggattgagattgtga 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	87.6	5518	6	AX281276
2	18.4	87.6	5518	6	AX344611
3	18.4	87.6	5518	6	AX345119
4	18.4	87.6	135372	10	AC103828
5	18.4	87.6	182686	10	AC121986
6	18.4	87.6	186048	2	AC074190
7	18.4	87.6	188868	9	AC023232
8	18.4	87.6	214787	2	AC123853
9	18	85.7	70122	9	AC091938
10	18	85.7	179686	2	AC023987
11	17.8	84.8	4369	2	AC096637
12	17.8	84.8	13377	6	AX251229
13	17.8	84.8	13377	6	AX346564
14	17.8	84.8	20303	3	AC006749
15	17.8	84.8	98368	9	AL136453
16	17.8	84.8	115291	2	AC019814
17	17.8	84.8	119987	2	CNS08086
18	17.8	84.8	123701	9	AL139236
19	17.8	84.8	146790	2	AC079342
20	17.8	84.8	160650	3	AC010695
21	17.8	84.8	160826	9	AC096898
22	17.8	84.8	162402	3	AC010574
23	17.8	84.8	165379	2	AC069265
24	17.8	84.8	168506	9	AC007404
25	17.8	84.8	184375	2	AC026219
26	17.8	84.8	190855	2	AL844490
27	17.8	84.8	190964	10	AL671848
28	17.8	84.8	197121	2	AC025520
29	17.8	84.8	215233	2	AC116590
30	17.8	84.8	224468	2	AL596455
31	17.8	84.8	224790	2	AL683819
32	17.8	84.8	279075	3	AE003530
33	17.4	82.9	139885	9	AC009232
34	17.4	82.9	186709	2	AC127994
35	17.4	82.9	193277	10	AL596096
36	17	81.0	16914	6	AX251953
37	17	81.0	16914	6	AX348748
38	17	81.0	120310	2	AP004132
39	17	81.0	149807	2	AP004998
40	17	81.0	171705	9	AC104166
41	17	81.0	218652	9	AC096922
42	16.8	80.0	353	11	G50981
43	16.8	80.0	477	8	AY094031
44	16.8	80.0	639	8	GA040600
45	16.8	80.0	836	8	AY069896

## ALIGNMENTS

RESULT 1  
AX281276  
LOCUS  
DEFINITION  
Sequence 18 from Patent WO0177164.  
ACCESSION  
AX281276.1 GI:16608531  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct.  
artificial sequences.  
REFERENCE  
1  
AUTHORS  
TITLE  
JOURNAL  
Olek, A., Piepenbrock, C. and Berlin, K.  
Diagnosis of diseases associated with apoptosis  
Patent: WO 0177164-A 18 OCT-2001;  
Epigenomics AG (DE)

Pred. No. is the number of results predicted by chance to have a



US-09-911-904-331

## Query Match

Best Local Similarity 69.1%; Score 15.2; DB 9; Length 116;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTATATCTACATTCGAAC 20

DB 95 CGTATGCTACACTCGAAGC 114

## RESULT 14

US-09-974-300-2342/c  
; Sequence 2342, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Clausen, Randy M.  
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene  
; FILE REFERENCE: 10085.500-US  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2342  
; LENGTH: 810  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-974-300-2342

## Query Match

Best Local Similarity 69.1%; Score 15.2; DB 10; Length 810;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 TATATCTACATTCGAACGA 22

DB 742 TATGCTACATTCGAATGA 723

## RESULT 15

US-09-758-269-9/c  
; Sequence 9, Application US/09758269  
; Patent No. US20020104120A1  
; GENERAL INFORMATION:  
; APPLICANT: IUCHI, SATOSHI  
; APPLICANT: KOBAYASHI, MASATOMO  
; APPLICANT: SHINOZAKI, KAZUO  
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN  
; FILE REFERENCE: 3914-3  
; CURRENT APPLICATION NUMBER: US/09/758,269  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: JP 2001-003476  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: JP 2000-010056  
; PRIOR FILING DATE: 2000-01-13  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 1734  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1731)  
US-09-758-269-9

## Query Match

69.1%; Score 15.2; DB 10; Length 1734;

Best Local Similarity 85.0%; Pred. No. 1.4e+03;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GTATATCTACATTCGAACG 21

DB 1471 GTATATCTACCTTCGCAATG 1452

Search completed: July 8, 2003, 06:04:14  
Job time : 31.1402 secs

US-09-973-945A-25

Query Match 70.0%; Score 15.4; DB 9; Length 29;  
Best Local Similarity 94.1%; Pred. No. 7.2e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TATATCTACATTCGAAA 19  
|||||  
DB 10 TATATCTATATTCGAAA 26

RESULT 10

US-09-973-945A-26/C  
Sequence 26, Application US/09973945A  
Patent No. US20020169297A1  
GENERAL INFORMATION:  
APPLICANT: University of Kentucky Research Foundation  
APPLICANT: XIE, Mingtang  
TITLE OF INVENTION: Genetic Insulator for Preventing Influence By Another Gene  
FILE REFERENCE: 050229-0287  
CURRENT APPLICATION NUMBER: US/09/973,945A  
CURRENT FILING DATE: 2001-10-11  
PRIOR APPLICATION NUMBER: 60/241,735  
PRIOR FILING DATE: 2000-10-20  
NUMBER OF SEQ ID NOS: 37  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 26  
LENGTH: 29  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-973-945A-26

Query Match 70.0%; Score 15.4; DB 9; Length 29;  
Best Local Similarity 94.1%; Pred. No. 7.2e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TATATCTACATTCGAAA 19  
|||||  
DB 20 TATATCTATATTCGAAA 4

RESULT 11

US-08-781-986A-1583/C  
Sequence 1583, Application US/08781986A  
Publication No. US20030054436A1  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5255  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,986A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benson, Bob  
REGISTRATION NUMBER: 30,446  
REFERENCE/DOCKET NUMBER: PB248BP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 1583:  
SEQUENCE CHARACTERISTICS:

LENGTH: 348 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-781-986A-1583

Query Match 70.0%; Score 15.4; DB 7; Length 348;  
Best Local Similarity 94.1%; Pred. No. 9.6e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TATATCTACATTCGAAA 19  
|||||  
DB 206 TATATCTACATTCGAAA 190

RESULT 12

US-09-911-904-216  
Sequence 216, Application US/09911904  
Publication No. US20030096234A1  
GENERAL INFORMATION:  
APPLICANT: Fair, Spencer B.  
APPLICANT: Pickett, Gavin G.  
APPLICANT: Neft, Robin Eileen  
APPLICANT: Dunn, II, Robert Thomas  
TITLE OF INVENTION: CANINE TOXICITY GENES  
FILE REFERENCE: 400742000200  
CURRENT APPLICATION NUMBER: US/09/911,904  
CURRENT FILING DATE: 2002-04-09  
PRIOR APPLICATION NUMBER: US 60/220,057  
PRIOR FILING DATE: 2000-07-21  
NUMBER OF SEQ ID NOS: 386  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 216  
LENGTH: 116  
TYPE: DNA  
ORGANISM: Canis familiaris  
US-09-911-904-216

Query Match 69.1%; Score 15.2; DB 9; Length 116;  
Best Local Similarity 85.0%; Pred. No. 1e+03;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGTATATCTACATTCGAAC 20  
|||||  
DB 95 CGTATGCTACAGTCGAAC 114

RESULT 13

US-09-911-904-331  
Sequence 331, Application US/09911904  
Publication No. US20030096234A1  
GENERAL INFORMATION:  
APPLICANT: Fair, Spencer B.  
APPLICANT: Pickett, Gavin G.  
APPLICANT: Neft, Robin Eileen  
APPLICANT: Dunn, II, Robert Thomas  
TITLE OF INVENTION: CANINE TOXICITY GENES  
FILE REFERENCE: 400742000200  
CURRENT APPLICATION NUMBER: US/09/911,904  
CURRENT FILING DATE: 2002-04-09  
PRIOR APPLICATION NUMBER: US 60/220,057  
PRIOR FILING DATE: 2000-07-21  
NUMBER OF SEQ ID NOS: 386  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 331  
LENGTH: 116  
TYPE: DNA  
ORGANISM: Canis familiaris

```
RESULT 6
US-09-738-626-2486/c
; Sequence 2486, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2486
; LENGTH: 1899
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2486

Query Match          70.9%; Score 15.6; DB 9; Length 1899;
Best Local Similarity 81.8%; Pred. No. 9.4e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 CGTATATCTACATTCGAACGA 22
        ||||| 1111111111111111
DB      495 CGCATAGCCACATTCGAATGCA 474

RESULT 7
US-10-239-676-162/c
; Sequence 162, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239, 676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 162
; LENGTH: 15306
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

```
US-10-239-676-162

Query Match          70.9%; Score 15.6; DB 9; Length 15306;
Best Local Similarity 81.8%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 CGTATATCTACATTCGAACGA 22
        ||||| 1111111111111111
DB      8794 CGAATCTACATTTAAACGA 8773

RESULT 8
US-09-738-626-1
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match          70.9%; Score 15.6; DB 9; Length 3309400;
Best Local Similarity 81.8%; Pred. No. 4.6e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 CGTATATCTACATTCGAACGA 22
        ||||| 1111111111111111
DB      2400809 CGCATAGCCACATTCGAATGCA 2400830

RESULT 9
US-09-973-945A-25
; Sequence 25, Application US/0973945A
; Patent No. US20020169297A1
; GENERAL INFORMATION:
; APPLICANT: University of Kentucky Research Foundation
; APPLICANT: GEN, Susheng
; APPLICANT: XIE, Mingtang
; TITLE OF INVENTION: Genetic Insulator for Preventing Influence By Another Gene
; FILE REFERENCE: 050229-0287
; CURRENT APPLICATION NUMBER: US/09/973, 945A
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 60/241,735
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 29
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
```

APPLICANT: Hu, Weng  
APPLICANT: Tishkoff, Daniel  
APPLICANT: Zamudio, Carlos  
APPLICANT: Eroskin, Alexey M  
APPLICANT: Lemieux, Sebastien M  
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
FILE REFERENCE: 10182-018-999  
CURRENT APPLICATION NUMBER: US/10/128,714  
CURRENT FILING DATE: 2002-04-23  
PRIOR APPLICATION NUMBER: US 60/285,697  
PRIOR FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: US 60/287,066  
PRIOR FILING DATE: 2001-04-27  
PRIOR APPLICATION NUMBER: US 60/295,890  
PRIOR FILING DATE: 2001-06-05  
PRIOR APPLICATION NUMBER: US 60/303,899  
PRIOR FILING DATE: 2001-07-09  
PRIOR APPLICATION NUMBER: US 60/316,362  
PRIOR FILING DATE: 2001-08-31  
NUMBER OF SEQ ID NOS: 8603  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 168  
LENGTH: 2909  
TYPE: DNA  
ORGANISM: Aspergillus fumigatus  
US-10-128-714-168

Query Match 73.6%; Score 16.2; DB 9; Length 2909;  
Best Local Similarity 85.7%; Pred. No. 5.3e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 CGTATATCTACATTCGAACG 21  
Db 2061 CTTATGCTACATTCGAACG 2041

RESULT 3  
US-10-128-714-5168/c  
Sequence 5168, Application US/10128714  
Publication No. US20030119013A1  
GENERAL INFORMATION:  
APPLICANT: Jiang, Bo  
APPLICANT: Hu, Weng  
APPLICANT: Tishkoff, Daniel  
APPLICANT: Zamudio, Carlos  
APPLICANT: Eroskin, Alexey M  
APPLICANT: Lemieux, Sebastien M  
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
FILE REFERENCE: 10182-018-999  
CURRENT APPLICATION NUMBER: US/10/128,714  
CURRENT FILING DATE: 2002-04-23  
PRIOR APPLICATION NUMBER: US 60/285,697  
PRIOR FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: US 60/287,066  
PRIOR FILING DATE: 2001-04-27  
PRIOR APPLICATION NUMBER: US 60/295,890  
PRIOR FILING DATE: 2001-06-05  
PRIOR APPLICATION NUMBER: US 60/303,899  
PRIOR FILING DATE: 2001-07-09  
PRIOR APPLICATION NUMBER: US 60/316,362  
PRIOR FILING DATE: 2001-08-31  
NUMBER OF SEQ ID NOS: 8603  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 5168  
LENGTH: 3224  
TYPE: DNA  
ORGANISM: Aspergillus fumigatus  
US-10-128-714-5168

Query Match 73.6%; Score 16.2; DB 9; Length 3224;  
Best Local Similarity 85.7%; Pred. No. 5.4e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
OY 1 CGTATATCTACATTCGAACG 21  
Db 2376 CTTATGCTACATTCGAACG 2356

RESULT 4  
US-10-108-605-70/c  
Sequence 70, Application US/10108605  
Patent No. US20020160934A1  
GENERAL INFORMATION:  
APPLICANT: Broadus, Julie  
APPLICANT: Stam, Lynn  
APPLICANT: Bachmann, Jane  
APPLICANT: Kamdar, Kim  
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCOD  
FILE REFERENCE: 3113B  
CURRENT APPLICATION NUMBER: US/10/108,605  
CURRENT FILING DATE: 2002-03-27  
PRIOR APPLICATION NUMBER: US 09/761,142  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/176,418  
PRIOR FILING DATE: 2000-01-14  
NUMBER OF SEQ ID NOS: 361  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 70  
LENGTH: 11527  
TYPE: DNA  
ORGANISM: Drosophila melanogaster  
US-10-108-605-70

Query Match 73.6%; Score 16.2; DB 9; Length 11527;  
Best Local Similarity 85.7%; Pred. No. 6.2e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 CGTATATCTACATTCGAACG 21  
Db 6421 CGTTATCTCTACATTCGAATG 6401

RESULT 5  
US-09-974-300-1167  
Sequence 1167, Application US/09974300  
Patent No. US20020146721A1  
GENERAL INFORMATION:  
APPLICANT: Berka, Randy M.  
APPLICANT: Clausen, Ib Groth  
TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
FILE REFERENCE: 10085.500-US  
CURRENT APPLICATION NUMBER: US/09/974,300  
CURRENT FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: 09/680,598  
PRIOR FILING DATE: 2000-10-06  
PRIOR APPLICATION NUMBER: 60/279,526  
PRIOR FILING DATE: 2001-03-27  
NUMBER OF SEQ ID NOS: 8481  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1167  
LENGTH: 1356  
TYPE: DNA  
ORGANISM: Bacillus licheniformis  
US-09-974-300-1167

Query Match 70.9%; Score 15.6; DB 10; Length 1356;  
Best Local Similarity 81.8%; Pred. No. 9.1e+02;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 CGTATATCTACATTCGAACG 22  
Db 685 CGTATCTCTACATTCGAACG 706

RESULT 2  
US-10-128-714-168/C  
; Sequence 168, Application US/10128714  
; Publication No. US20030119013A1  
; GENERAL INFORMATION:  
; APPLICANT: JiaGu, Bo

QY 1 CGTATCTACATTCGAAACG 21  
 ||||||| ||||| |||||  
 Db 427 CGCATATCAACTTCAAATG 447

RESULT 15  
 US-09-605-785-633/c  
 ; Sequence 633, Application US/09605785  
 ; Patent No. 6321716  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Dillon, Davin C.  
 ; APPLICANT: Mitcham, Jennifer L.  
 ; APPLICANT: Harlocker, Susan L.  
 ; APPLICANT: Jiang, Yugu  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Kalos, Michael D.  
 ; APPLICANT: Fanger, Gary R.  
 ; APPLICANT: Retter, Marc W.  
 ; APPLICANT: Stolk, John A.  
 ; APPLICANT: Day, Craig H.  
 ; APPLICANT: Vedwick, Thomas S.  
 ; APPLICANT: Carter, Darrick  
 ; APPLICANT: Li, Samuel  
 ; APPLICANT: Wang, Aijun  
 ; APPLICANT: Skeiky, Yasir A.W.  
 ; APPLICANT: Hepler, William  
 ; APPLICANT: Hepler, William  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 ; FILE REFERENCE: 210121.427C16  
 ; CURRENT APPLICATION NUMBER: US/09/605,785  
 ; CURRENT FILING DATE: 2000-06-27  
 ; NUMBER OF SEQ ID NOS: 835  
 ; SOFTWARE: FASTSEQ for Windows Version 3.0  
 ; SEQ ID NO 633  
 ; LENGTH: 630  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(630)  
 ; OTHER INFORMATION: n = A,T,C or G  
 ; US-09-605-785-633

Query Match 66.4%; Score 14.6; DB 4; Length 630;  
 Best local Similarity 81.0%; Pred. No. 1.6e+02;  
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GTATATCTACATTCGAAACG 22  
 ||||||| ||||| |||||  
 Db 366 GTATATCTACATTCCTATGA 346

Search completed: July 8, 2003, 05:56:44  
 Job time : 27.2769 secs

OY 1 CGTATATCTACATTCGAAAG 21  
|| ||||| ||||| |||||  
Db 427 CGCATATCAACATTCAAAATG 447

## RESULT 12

US-08-779-379-23  
Sequence 23, Application US/08779379  
Patent No. 5858970  
GENERAL INFORMATION:  
APPLICANT: Zeelon, Elisha P.  
APPLICANT: Weiber, Moshe M.  
APPLICANT: Levanon, Avigdor  
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/779,379  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/225,442  
FILING DATE: 08-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0317/43020-A/JPW/EAB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-977-9550  
TELEFAX: 212-664-0525  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 469 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
US-08-779-379-23

Query Match 66.4%; Score 14.6; DB 2; Length 469;  
Best Local Similarity 81.0%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 CGTATATCTACATTCGAAAG 21  
|| ||||| ||||| |||||  
Db 427 CGCATATCAACATTCAAAATG 447

RESULT 13  
US-08-469-219-23  
Sequence 23, Application US/08469219  
Patent No. 5863534  
GENERAL INFORMATION:  
APPLICANT: Zeelon, Elisha P.  
APPLICANT: Weiber, Moshe M.  
APPLICANT: Levanon, Avigdor  
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,219  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/225,442  
FILING DATE: 08-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0317/43020-A/JPW/EAB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-977-9550  
TELEFAX: 212-664-0525  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 469 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
US-08-469-219-23

Query Match 66.4%; Score 14.6; DB 2; Length 469;  
Best Local Similarity 81.0%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 CGTATATCTACATTCGAAAG 21  
|| ||||| ||||| |||||  
Db 427 CGCATATCAACATTCAAAATG 447

RESULT 14  
US-09-228-152-23  
Sequence 23, Application US/09228152  
Patent No. 6211341  
GENERAL INFORMATION:  
APPLICANT: Zeelon, Elisha P.  
APPLICANT: Weiber, Moshe M.  
APPLICANT: Levanon, Avigdor  
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa INHIBITORY ACTIVITY  
FILE REFERENCE: 43020A2A  
CURRENT APPLICATION NUMBER: US/09/228,152  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 23  
LENGTH: 469  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: DNA sequence  
US-09-228-152-23

Query Match 66.4%; Score 14.6; DB 4; Length 469;  
Best Local Similarity 81.0%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,219  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/225,442  
FILING DATE: 08-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0317/43020-A/JPW/EAB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-977-9550  
TELEFAX: 212-664-0525  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 469 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
US-08-469-219-23

Query Match 66.4%; Score 14.6; DB 2; Length 469;  
Best Local Similarity 81.0%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 CGTATATCTACATTCGAAAG 21  
|| ||||| ||||| |||||  
Db 427 CGCATATCAACATTCAAAATG 447

RESULT 14  
US-09-228-152-23  
Sequence 23, Application US/09228152  
Patent No. 6211341  
GENERAL INFORMATION:  
APPLICANT: Zeelon, Elisha P.  
APPLICANT: Weiber, Moshe M.  
APPLICANT: Levanon, Avigdor  
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa INHIBITORY ACTIVITY  
FILE REFERENCE: 43020A2A  
CURRENT APPLICATION NUMBER: US/09/228,152  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 23  
LENGTH: 469  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: DNA sequence  
US-09-228-152-23

Query Match 66.4%; Score 14.6; DB 4; Length 469;  
Best Local Similarity 81.0%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

APPLICATION NUMBER: US/08/468,347  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/225,442  
FILING DATE: 08-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0317/43020-A/JPW/EAB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-977-9550  
TELEFAX: 212-664-0525  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 469 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
US-08-468-347-23

Query Match  
Best Local Similarity 81.0%; Score 14.6; DB 1; Length 469;  
Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGTATCTACTTCGAAACG 21  
Db 427 CGCATATCAACATTCAAATG 447

RESULT 10  
US-08-226-264-25  
Sequence 25, Application US/08226264  
Patent No. 5801017  
GENERAL INFORMATION:  
APPLICANT: Weiber, Moshe M.  
APPLICANT: Zeelon, Elisha P.  
APPLICANT: Levanon, Avigdor  
APPLICANT: Guy, Rachel  
APPLICANT: Goldlust, Arie  
APPLICANT: Rigbi, Meir  
APPLICANT: Panet, Amos  
APPLICANT: Fischer, Meir  
TITLE OF INVENTION: PRODUCTION OF RECOMBINANT FACTOR XA  
TITLE OF INVENTION: INHIBITORS OF LEECH HIRUDO MEDICINALIS  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/226,264  
FILING DATE: 08-APR-94  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 40017-A/JPW/G3G  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525

TELEX:  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 469 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-226-264-25

Query Match  
Best Local Similarity 81.0%; Score 14.6; DB 1; Length 469;  
Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGTATCTACTTCGAAACG 21  
Db 427 CGCATATCAACATTCAAATG 447

RESULT 11  
US-08-467-389-23  
Sequence 23, Application US/08467389  
Patent No. 5824641  
GENERAL INFORMATION:  
APPLICANT: Zeelon, Elisha P.  
APPLICANT: Weiber, Moshe M.  
APPLICANT: Levanon, Avigdor  
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa  
TITLE OF INVENTION: INHIBITORY ACTIVITY  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,389  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/225,442  
FILING DATE: 08-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0317/43020-A/JPW/EAB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-977-9550  
TELEFAX: 212-664-0525  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 469 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
US-08-467-389-23

Query Match  
Best Local Similarity 81.0%; Score 14.6; DB 1; Length 469;  
Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;



Patent No. 6030832  
GENERAL INFORMATION:  
APPLICANT: Mong, Alexander K.C.  
APPLICANT: Bartel, Paul L.  
APPLICANT: Teng, David H.-F.  
APPLICANT: Tavligian, Sean V.  
TITLE OF INVENTION: A Carboxy-Terminal BRCA1 Interacting  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.  
STREET: 555 Thirteenth Street, N.W., Suite 701 East  
CITY: Washington  
STATE: DC  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/975,703  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Saxe, Stephen A.  
REGISTRATION NUMBER: 38,609  
REFERENCE/DOCKET NUMBER: 2318-0174  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-624-1589  
TELEFAX: 202-783-6031  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2694 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2691  
US-08-975-703-5

Query Match 67.3%; Score 14.8; DB 3; Length 2694;  
Best Local Similarity 88.9%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSTATCTACATTCGAAA 19  
DB 1103 GSTATCTACATTCGAAA 1120

RESULT 8  
US-09-515-884-5  
Sequence 5, Application US/09515884  
Patent No. 6235263  
GENERAL INFORMATION:  
APPLICANT: Mong, Alexander K.C.  
APPLICANT: Bartel, Paul L.  
APPLICANT: Teng, David H.-F.  
APPLICANT: Tavligian, Sean V.  
TITLE OF INVENTION: A Carboxy-Terminal BRCA1 Interacting  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.  
STREET: 555 Thirteenth Street, N.W., Suite 701 East

CITY: Washington  
STATE: DC  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: 08/975,703  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Saxe, Stephen A.  
REGISTRATION NUMBER: 38,609  
REFERENCE/DOCKET NUMBER: 2318-0174  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-624-1589  
TELEFAX: 202-783-6031  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2694 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2691  
US-09-515-884-5

Query Match 67.3%; Score 14.8; DB 4; Length 2694;  
Best Local Similarity 88.9%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSTATCTACATTCGAAA 19  
DB 1103 GSTATCTACATTCGAAA 1120

RESULT 9  
US-08-468-347-23  
Sequence 23, Application US/08468347  
Patent No. 5783421  
GENERAL INFORMATION:  
APPLICANT: Zeeion, Elisha P.  
APPLICANT: Werber, Moshe M.  
APPLICANT: Levanon, Avigdor  
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 1034:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2018 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORPHYROMONAS GINGIVALIS  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1...2018  
US-09-221-017B-1034

Query Match 67.3% Score 14.8; DB 4; Length 2018;  
Best Local Similarity 88.9%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 TATCTACATTCGAACGA 22  
|||||  
Db 1169 TATCGACATTCGATACGA 1152

RESULT 5  
US-09-134-078-11  
Sequence 11, Application US/09134078  
Patent No. 6368844  
GENERAL INFORMATION:  
APPLICANT: Bylina, Edward J.  
TITLE OF INVENTION: GLYCOSIDASE ENZYMES  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gray Cary Ware & Freidenrich LLP  
STREET: 4365 Executive Drive, Suite 1600  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/134,078  
FILING DATE: 13-AUG-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/949,026  
FILING DATE: 10-OCT-1997  
APPLICATION NUMBER: 60/056,916  
FILING DATE: 06-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/024002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 858/677-1456  
TELEFAX: 858/677-1465  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2043 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA

FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1...2040  
US-09-134-078-11

Query Match 67.3% Score 14.8; DB 4; Length 2043;  
Best Local Similarity 88.9%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATATCTACATTCGAACG 21  
|||||  
Db 1754 ACATCTACATTCGACAA 1771

RESULT 6  
US-08-914-999-7/C  
Sequence 7, Application US/08914999  
Patent No. 6346406  
GENERAL INFORMATION:  
APPLICANT: Ryzanov, Alexey G.  
APPLICANT: Pavur, Karen S.  
TITLE OF INVENTION: ELONGATION FACTOR-2 KINASE (EF-2 KINASE)  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David A. Jackson, Esq.  
STREET: 411 Hackensack Ave, Continental Plaza, 4th  
STREET: Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/914,999  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 601-1-078  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2237 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Dictyostelium discoideum  
US-08-914-999-7

Query Match 67.3% Score 14.8; DB 4; Length 2237;  
Best Local Similarity 88.9%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GATATCTACATTCGAA 19  
|||||  
Db 1745 GATATTTACATTCCAA 1728

RESULT 7  
US-08-975-703-5  
Sequence 5, Application US/08975703

SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221,017B  
FILING DATE: 23-DEC-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1182  
FILING DATE: 31-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1546  
FILING DATE: 30-JAN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP2911  
FILING DATE: 09-APR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU98/01023  
FILING DATE: 10-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Montoy, Gladys H  
REGISTRATION NUMBER: 32,430  
REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4212 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORPHYROMOMAS GINGIVALIS  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1...4212  
US-09-221-017B-39

Query Match 69.1%; Score 15.2; DB 4; Length 4212;  
Best Local Similarity 85.0%; Pred. No. 95;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 TATATCTACATTCGAACGA 22  
||||| ||||| |||||  
Db 2978 TATATATAATTAGAACGA 2959

RESULT 3  
US-09-134-078-57  
Sequence 57, Application US/09134078  
Patent No. 6368844  
GENERAL INFORMATION:  
APPLICANT: Bylina, Edward J.  
TITLE OF INVENTION: GIXCOSIDASE ENZYMES  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gray Cary Ware & Freidenrich LLP  
STREET: 4365 Executive Drive, Suite 1600  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/134,078  
FILING DATE: 13-AUG-1998

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/349,026  
FILING DATE: 10-OCT-1997  
APPLICATION NUMBER: 60/056,916  
FILING DATE: 06-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/024002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 858/677-1456  
TELEFAX: 858/677-1465  
INFORMATION FOR SEQ ID NO: 57:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1992 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
MOLECULE TYPE: Genomic DNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1...1983  
US-09-134-078-57

Query Match 67.3%; Score 14.8; DB 4; Length 1992;  
Best Local Similarity 88.9%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATATCTACATTCGAACG 21  
||||| ||||| |||||  
Db 1703 ACATCTACATTCGAACG 1720

RESULT 4  
US-09-221-017B-1034/C  
Sequence 1034, Application US/09221017B  
Patent No. 6444799  
GENERAL INFORMATION:  
APPLICANT: Ross, Bruce C.  
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
NUMBER OF SEQUENCES: 1120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221,017B  
FILING DATE: 23-DEC-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1182  
FILING DATE: 31-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1546  
FILING DATE: 30-JAN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP2911  
FILING DATE: 09-APR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU98/01023  
FILING DATE: 10-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Montoy, Gladys H  
REGISTRATION NUMBER: 32,430

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:45:20 ; Search time 5.27694 Seconds  
(without alignments)  
1278.561 Million cell updates/sec

Title: US-09-477-082-30

Perfect score: 22

Sequence: 1 cgatattacattcgaaacga 22

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 segs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database : Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCRTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfilest.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	15.2	69.1	999	4	US-09-134-001C-1178
C 2	15.2	69.1	4212	4	US-09-221-017B-39
C 3	14.8	67.3	1992	4	US-09-134-078-57
C 4	14.8	67.3	2018	4	US-09-221-017B-1034
C 5	14.8	67.3	2043	4	US-09-134-078-11
C 6	14.8	67.3	2237	4	US-08-914-999-7
C 7	14.8	67.3	2694	3	US-08-975-703-5
C 8	14.8	67.3	2694	4	US-09-515-884-5
C 9	14.6	66.4	469	1	US-08-468-347-23
C 10	14.6	66.4	469	1	US-08-226-264-25
C 11	14.6	66.4	469	1	US-08-467-389-23
C 12	14.6	66.4	469	2	US-08-779-379-23
C 13	14.6	66.4	469	2	US-08-469-219-23
C 14	14.6	66.4	469	4	US-09-228-152-23
C 15	14.6	66.4	630	4	US-09-605-785-633
C 16	14.6	66.4	780	4	US-09-134-001C-1631
C 17	14.6	66.4	1410	2	US-08-975-316-86
C 18	14.6	66.4	1410	4	US-09-615-192A-86
C 19	14.6	66.4	2112	4	US-09-232-160-12
C 20	14.6	66.4	2728	4	US-09-232-160-1
C 21	14.6	66.4	3504	4	US-08-857-076-47
C 22	14.6	66.4	3959	1	US-08-474-067-1
C 23	14.6	66.4	3959	2	US-08-474-068A-1
C 24	14.6	66.4	11811	4	US-09-078-294-7
C 25	14.6	65.5	369	1	US-09-053-945-3
C 26	14.4	65.5	607	1	US-09-053-945-1
C 27	14.4	65.5	607	1	US-09-053-945-1

C 28	14.4	65.5	607	1	US-08-061-314A-1	Sequence 1, Appl
C 29	14.4	65.5	711	4	US-09-134-001C-1834	Sequence 1834, Ap
C 30	14.4	65.5	2661	4	US-09-134-001C-1659	Sequence 1659, Ap
C 31	14.4	65.5	5055	4	US-09-242-632A-13	Sequence 13, Appl
C 32	14.4	65.5	31491	4	US-09-360-186-1	Sequence 1, Appl
C 33	14.2	64.5	213	4	US-09-134-001C-1113	Sequence 1113, Ap
C 34	14.2	64.5	222	4	US-09-134-001C-599	Sequence 599, App
C 35	14.2	64.5	470	4	US-09-615-192A-189	Sequence 189, App
C 36	14.2	64.5	609	1	US-08-268-072-2	Sequence 2, Appl
C 37	14.2	64.5	609	1	US-08-690-721-2	Sequence 2, Appl
C 38	14.2	64.5	684	1	US-08-226-264-27	Sequence 27, Appl
C 39	14.2	64.5	836	4	US-09-221-017B-878	Sequence 878, App
C 40	14.2	64.5	1314	4	US-09-134-001C-581	Sequence 581, App
C 41	14.2	64.5	1521	4	US-09-134-001C-470	Sequence 470, App
C 42	14.2	64.5	1812	2	US-08-669-536-1	Sequence 1, Appl
C 43	14.2	64.5	2000	1	US-09-041-075A-1	Sequence 1, Appl
C 44	14.2	64.5	3030	4	US-09-693-146-3	Sequence 3, Appl
C 45	14.2	64.5	3927	4	US-09-293-238B-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-09-134-001C-1178/c  
Sequence 1178, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
PRIOR FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 1178  
LENGTH: 999  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-1178

Query Match 69.1%; Score 15.2; DB 4; Length 999;  
Best Local Similarity 85.0%; Pred. No. 87;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Caps 0;

QY 3 TATATCTACATTCGAACGA 22  
DB 508 TATATCTACATTCGAACGA 489

RESULT 2  
US-09-221-017B-39/c  
Sequence 39, Application US/09221017B  
Patent No. 6444799  
GENERAL INFORMATION:  
APPLICANT: Ross, Bruce C.  
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
NUMBER OF SEQUENCES: 1120  
CORRESPONDENCE ADDRESS:  
ADDRESS: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows

TITLE  
P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,  
Church, S., Jackson, L. and Bradford, K.  
Lettuce and Sunflower ESTs from the Compositae Genome Project  
<http://compenomics.ucdavis.edu/>  
Unpublished (2002)

JOURNAL  
COMMENT  
Contact: Alexander Kozik [R.W.Michelmore]  
Department of Vegetable Crops, R.W.Michelmore Lab  
University of California at Davis (UCD)  
Asmundson Hall, UCD, Davis, CA 95616, USA  
Tel: 1-(530)-742-1742  
Fax: 1-(530)-752-9659  
Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]  
singleton, see <http://cgpbdb.ucdavis.edu/> for details.  
Plate: QG16 row: M column: 15.  
Location/Qualifiers

## FEATURES

source

1. 599  
/organism="Lactuca sativa"  
/cultivar="Salinas"  
/db\_xref="taxon:4236"  
/clone="QG16M15"  
/lab\_host="E.coli"  
/lab\_lib="QG-ABCDI lettuce salinas"  
/note="Vector: pBRCDNA5flab: The library was constructed  
from 10 different sources of RNA from a single genotype.  
Separate cDNAs were generated using primers that  
incorporated unique 5' and 3' tags to distinguish each  
source of RNA. cDNAs were then pooled, size-fractionated,  
directionally cloned into a custom medium-copy vector and  
transformations made with four size classes to minimize  
size bias. Details of each source of RNA and library  
construction can be obtained at <http://cgpbdb.ucdavis.edu/>  
TAG-LIB-QG-ABCDI lettuce salinas  
TAG-TISSUE=roots  
TAG\_SEQ=GTTCACGCGC"

## BASE COUNT

201 a 113 c 121 g 164 t

## ORIGIN

Query Match 76.4%; Score 16.8; DB 14; Length 599;  
Best Local Similarity 90.0%; Pred. No. 4.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TATATCTACATTCGAACGA 22  
||||| ||||||| ||  
Db 440 TATATCTACATTCGAACGA 459

## RESULT 15

AQ311480

LOCUS

CITR1-EL-2526L19.TR

639 bp DNA

linear

GSS 22-DEC-1998

## DEFINITION

DNA sequence.

## ACCESSION

AQ311480

VERSION

AQ311480.1

GI:4039833

## KEYWORDS

GSS.

## SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

## REFERENCE

1 (bases 1 to 639)

Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,  
Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and  
Venter, J.C.

## AUTHORS

Use of a random human BAC End Sequence Database for Sequence-Ready

## TITLE

Map Building (1998)

## JOURNAL

Unpublished (1998)

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbeet19f.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:  
[http://www.tigr.org/tcbl/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tcbl/humgen/bac_end_search/bac_end_search.html).  
Seq primer: M13 Reverse  
Class: BAC ends.

## FEATURES

source

Location/Qualifiers  
1. 639  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2526L19"  
/clone\_lib="CITR1-EL"  
/sex="male"  
/cell\_type="sperm"  
/note="Vector: pBeloBAC11, Site\_1: EcoRI, Site\_2: EcoRI;  
Caltech Human BAC Library D"

## BASE COUNT

205 a 122 c 142 g 170 t

## ORIGIN

Query Match 76.4%; Score 16.8; DB 17; Length 639;  
Best Local Similarity 90.0%; Pred. No. 4.7e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TATATCTACATTCGAACGA 22  
||||| ||||||| ||  
Db 206 TATATCTATATTCGAACGA 225

Search completed: July 8, 2003, 05:51:22  
Job time: 170.879 secs

FEATURES Seq primer: M13uni primer for 3' end.  
Location/Qualifiers  
Source 1..3324  
/organism="Hordeum vulgare"  
/cultivar="Barke"  
/db\_xref="taxon:4513"  
/clone\_1lb="HM04613u"  
/clone\_1lb="Hordeum vulgare Barke roots"  
/tissue\_type="roots"  
/lab\_host="X10LR"  
/note="Vector: plasmid pBK-CMV; Site\_1: EcoRI; Site\_2: XhoI; mRNA was made from roots of spring barley variety 'Barke', a high quality malting variety. Roots were grown for two days on filter paper at room temperature. Cloning sites: EcoRI (5'-end of cDNA) and XhoI (3'-end of cDNA).  
NOTE: Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter. Average insert size is 1 kb. Sequence trimming: Vector sequences and sequence ends were trimmed from the 5'- and 3'-end until a 50 bp window contains less than two ambiguities. The maximum length was set to 700 bp"

BASE COUNT 73 a 89 c 49 g 108 t 5 others  
ORIGIN

Query Match 76.4%; Score 16.8; DB 9; Length 324;  
Best Local Similarity 90.0%; Pred. No. 3.7e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGTATCTCATCTCGAAC 20  
157 CGTATCTCATCTCGAAC 176

RESULT 12  
LOCUS AU037723 451 bp mRNA linear EST 29-MAR-1999  
DEFINITION AU037723 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium  
discoidium cDNA clone SSE216, mRNA sequence.  
ACCESSION AU037723  
VERSION AU037723.1 GI:3984476  
KEYWORDS EST.  
SOURCE Dictyostelium discoideum.  
ORGANISM Dictyostelium discoideum.  
REFERENCE 1 (bases 1 to 451)  
Mori, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M., Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.  
Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,  
The Dictyostelium developmental cDNA project: generation and  
analysis of expressed sequence tags from the first-finger stage of  
development  
JOURNAL DNA Res. 5 (6), 335-340 (1998)  
MEDLINE 99156227  
COMMENT Contact: Hideko Urushihara  
Institute of Biological Sciences  
University of Tsukuba  
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan  
Tel: 81-298-53-4664  
Fax: 81-298-53-6614  
Email: hideko@biol.tsukuba.ac.jp  
PROJECT = "Dictyostelium discoideum cDNA project in Japan".

FEATURES  
Source  
1..451  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone\_1lb="SSE216"  
/clone\_1lb="Dictyostelium discoideum SS (H.Urushihara)"  
/dev\_stage="slug"  
BASE COUNT 170 a 44 c 82 g 155 t  
ORIGIN

Query Match 76.4%; Score 16.8; DB 9; Length 451;  
Best Local Similarity 90.0%; Pred. No. 3.7e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Best Local Similarity 90.0%; Pred. No. 4.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 TATATCTACATTCGAACGA 22  
335 TATATCTACATTCGAACGA 354

RESULT 13  
LOCUS AU037770 452 bp mRNA linear EST 29-MAR-1999  
DEFINITION AU037770 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium  
discoidium cDNA clone SSE272, mRNA sequence.  
ACCESSION AU037770  
VERSION AU037770.1 GI:3984523  
KEYWORDS EST.  
SOURCE Dictyostelium discoideum.  
ORGANISM Dictyostelium discoideum.  
REFERENCE 1 (bases 1 to 452)  
Mori, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,  
Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,  
Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.  
The Dictyostelium developmental cDNA project: generation and  
analysis of expressed sequence tags from the first-finger stage of  
development  
JOURNAL DNA Res. 5 (6), 335-340 (1998)  
MEDLINE 99156227  
COMMENT Contact: Hideko Urushihara  
Institute of Biological Sciences  
University of Tsukuba  
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan  
Tel: 81-298-53-4664  
Fax: 81-298-53-6614  
Email: hideko@biol.tsukuba.ac.jp  
PROJECT = "Dictyostelium discoideum cDNA project in Japan".

FEATURES  
source  
1..452  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone\_1lb="SSE272"  
/clone\_1lb="Dictyostelium discoideum SS (H.Urushihara)"  
/dev\_stage="slug"  
BASE COUNT 170 a 45 c 82 g 155 t  
ORIGIN

Query Match 76.4%; Score 16.8; DB 9; Length 452;  
Best Local Similarity 90.0%; Pred. No. 4.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 TATATCTACATTCGAACGA 22  
336 TATATCTACATTCGAACGA 355

RESULT 14  
LOCUS B0874972 599 bp mRNA linear EST 15-AUG-2002  
DEFINITION B0874972 OGI015.yg.abi OG ABCDI lettuce salinas lactuca sativa cDNA clone  
OGI015, mRNA sequence.  
ACCESSION B0874972  
VERSION B0874972.1 GI:22261532  
KEYWORDS EST.  
SOURCE Lactuca sativa.  
ORGANISM Lactuca sativa.  
REFERENCE 1 (bases 1 to 599)  
Kozlik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,  
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison

**AUTHORS**  
Unushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.

**TITLE**  
Full length cDNA of Dictyostelium discoideum at the culmination stage

**JOURNAL**  
Unpublished (2002)

**COMMENT**  
Contact: Tadasu Shin-1  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8640, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp.  
Location/Qualifiers

**FEATURES**  
**source**  
1. .136  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="ddc35g24"  
/clone\_1b="Dictyostelium discoideum cDNA library, CF"  
/sex="mat A"  
/dev\_stage="Culmination stage"

**BASE COUNT**  
47 a 12 c 11 g 66 t

**ORIGIN**

**Query Match**  
Best Local Similarity 76.4%; Score 16.8; DB 13; Length 136;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

**Oy**  
3 TATATCATTCGACGAACA 22  
||||| ||||| ||||| ||  
95 TATATCATTCGACGAACA 76

**Db**

**RESULT 10**  
BB310076  
LOCUS BB310076/c

**DEFINITION**  
BB310076 RIKEN full-length enriched, adult male corpora quadrigenita Mus musculus cDNA clone B23031AA20 3' similar to D85189 Rattus norvegicus mRNA for Acyl-CoA synthetase, mRNA sequence.

**VERSION**  
BB310076

**KEYWORDS**  
BB310076.1 GI:9010781

**SOURCE**  
EST.

**ORGANISM**  
house mouse.  
Mus musculus

**REFERENCE**  
Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus. 1 (bases 1 to 204)  
Kono, H., Alzawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Itzawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kituchi, N., Kiyoawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

**COMMENTS**  
RIKEN Mouse ESTs (Kono, H., et al.)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22, Suhiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resgsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasakhi, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermostabilization and Thermoinactivation of thermolabile enzymes by

	trehalose and its application for the synthesis of full length cDNA. Proc Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)					
	Itoh,M., Kitsuai,T., Akiyama,O., Shibata,K., Iwasa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.					
	Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)					
	Carninci,P. and Hayashizaki,Y.					
	High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)					
	Please visit our web site ( <a href="http://genome.rtc.riken.go.jp">http://genome.rtc.riken.go.jp</a> ) for further details.					
FEATURES	Location/Qualifiers					
source	1..204					
	/organism="Mus musculus"					
	/db_xref="taxon:10090"					
	/clone="B230314A20"					
	/clone_id="RIKEN full-length enriched, adult male corpora quadrigemina"					
	/sex="male"					
	/tissue_type="corpora quadrigemina"					
	/dev_stage="adult"					
	/lab_host="DH10B"					
	/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGAGAGAGCATCCACAGCTCTTTTTTTTTTTNN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGATTCGCAGATTAAATAATATATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified Bluescript KS(+) after bulk excision from Lambda FLC I."					
BASE COUNT	55 a 35 c 28 g 86 t					
ORIGIN						
Query Match	76.4%; Score 16.8; DB 10; Length 204;					
Best Local Similarity	90.0%; Freq. No. 3.2e+02;					
Matches	18; Conservative 0; Mismatches 2; Indels 0; Gaps 0.					
CY	2 STATATCTACATCGAACG 21 					
Dd	56 GTATACTACTGTCAATG 37					
RESULT 11						
AL501502	324 bp mRNA linear EST 04-JAN-2001					
LOCUS	AL501502 Hordeum vulgare Barke roots Hordeum vulgare cDNA clone					
DEFINITION	HM04G13u 3', mRNA sequence.					
ACCESSION	AL501502					
VERSION	AL501502.1 GI:12027717					
KEYWORDS	EST.					
SOURCE	Hordeum vulgare.					
ORGANISM	Hordeum vulgare					
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.					
	1 (bases 1 to 324)					
REFERENCE	Michalek,W., Weschke,W., Pleissner,K.-P. and Graner,A.					
AUTHORS	EST sequencing and analysis in barley					
TITLE	Unpublished (2000)					
JOURNAL	Contact: Michalek W					
COMMENT	Institute for Plant Genetics and Crop Plant Research Corrensstr.3 D-06466 Gatersleben, Germany Email: michalek@ipk-gatersleben.de, <a href="http://pgrc.ipk-gatersleben.de">http://pgrc.ipk-gatersleben.de</a>					

JOURNAL  
REFERENCE 2 (bases 1 to 660)  
AUTHORS  
TITLE  
JOURNAL  
Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
- Web : www.genoscope.cns.fr

FEATURES  
source  
1. 660  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="T9J4"  
/clone\_lib="TAKU"  
/note="end : SP6"

BASE COUNT  
202 a 148 c 133 g 177 t

ORIGIN  
Query Match 78.2%; Score 17.2; DB 17; Length 660;  
Best Local Similarity 86.4%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  
1 CGTATCTACATTCGAAACGA 22  
|||||  
554 CGAGATCTACATTCGATGGA 533

RESULT 7  
BH530096/c 817 bp DNA linear GSS 14-DEC-2001  
LOCUS  
DEFINITION  
BOHQ091R BOHQ Brassica oleracea genomic clone BOHQ09, DNA  
sequence.  
ACCESSION  
BH530096  
VERSION  
BH530096.1 GI:17745951  
KEYWORDS  
GSS.  
SOURCE  
Brassica oleracea.  
ORGANISM  
Brassica oleracea.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
1 (bases 1 to 817)  
Town, C.D., Van Aken, S., Uteback, T. and Fraser, C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other\_GSSs: BOHQ091R  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: FR  
Class: sheared ends.  
Location/Qualifiers  
1. 817  
/organism="Brassica oleracea"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOHQ09"  
/clone\_lib="BOHQ"  
/note="Vector: PHOSI; Site: 1; BstXI; 2-3 kb sheared  
genomic DNA inserted into PHOSI using BstXI linkers"

BASE COUNT  
245 a 203 c 180 g 189 t

ORIGIN  
Query Match 78.2%; Score 17.2; DB 17; Length 817;  
Best Local Similarity 86.4%; Pred. No. 3.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  
1 CGTATCTACATTCGAAACGA 22  
|||||

DB 279 CTTATCTCGCATTTGAAACGA 258

RESULT 8  
BU156899  
LOCUS  
DEFINITION  
BU156899 full length cDNA library, chloronemata and young  
gametophores Physcomitrella patens subsp. patens cDNA clone pph2118  
5', mRNA sequence.

ACCESSION  
BU156899  
VERSION  
BU156899.1 GI:18324904  
KEYWORDS  
SOURCE  
ORGANISM  
Physcomitrella patens subsp. patens.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;  
Bryopsida; Funariidae; Funariaceae; Physcomitrella.  
1 (bases 1 to 587)  
Fujita, T., Shin-I, T., Seki, M., Kamiya, A., Uchiyama, T., Nishiyama, T.,  
Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y. and Hasebe  
/M.

REFERENCE  
Comparison of the moss Physcomitrella patens genome with flowering  
plants genome  
Unpublished (2002)  
Contact: Tadasu Shin-I  
Center for Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tsuhin@genes.nig.ac.jp

A backbone of the vector is basically from pBluescript(ks), that  
was in vivo excised from a modified lps phage vector (Mo bi Rec,  
Germany). 5' end of the cDNA that was digested with XhoI was  
ligated to SalI site of the vector and the 3' end including polyA  
tail was ligated to BamHI site of the vector. cDNA insert could be  
amplified with conventional 17 and 13 primers. This full-length  
cDNA library was generated basically according to the method  
described in The Plant J 15, 707-720 (1998) Seki M. et al.  
Piconemata were blended by the POLYTRON, and then cultivated on  
the BODAG medium for 13-14 days under the continuous light.

FEATURES  
source  
1. 587  
/organism="Physcomitrella patens subsp. patens"  
/db\_xref="taxon:145481"  
/clone="pph2118"  
/clone\_lib="full length cDNA library, chloronemata and  
young gametophores"  
/tissue-type="mixture of chloronemata and young  
gametophores with 2 to 5 leaves"

BASE COUNT  
126 a 132 c 202 g 127 t

ORIGIN  
Query Match 77.3%; Score 17; DB 13; Length 587;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  
1 CGTATCTACATTCGA 17  
|||||

DB 569 CGTATCTACATTCGA 565

RESULT 9  
BU379738/c 136 bp mRNA linear EST 08-MAR-2002  
LOCUS  
DEFINITION  
BU379738 Dictyostelium discoideum cDNA library, CF Dictyostelium  
discoideum cDNA clone ddc35g24 3', mRNA sequence.

ACCESSION  
BU379738  
VERSION  
BU379738.1 GI:19289121  
KEYWORDS  
SOURCE  
ORGANISM  
Dictyostelium discoideum.  
Dictyostelium discoideum  
Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
1 (bases 1 to 136)



COMMENT Tel:81-45-503-9111, Fax:81-45-503-9170  
Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

## PRIMERS

Sequencing: M13Rev  
LIBRARY Vector : pKS145  
R.Site 1 : SacI  
R.Site 2 : SacI

FEATURES Location/Qualifiers  
1..1970

/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-099107.R"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC Library"

BASE COUNT 364 a 405 c 473 g 576 t 132 others  
ORIGIN

Query Match 80.9%; Score 17.8; DB 17; Length 1970;  
Best Local Similarity 86.4%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 CGTATCTACATCGAAGCA 22  
Db 1746 CATATCTACATCGAAGCA 1725

RESULT 4  
LOCUS BH582702 791 bp DNA linear GSS 15-DEC-2001  
DEFINITION B0GTC24TR B0GT Brassica oleracea genomic clone B0GTC24, DNA sequence.

ACCESSION BH582702  
VERSION BH582702.1 GI:17835159  
KEYWORDS GSS

SOURCE Brassica oleracea.  
ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 791)

REFERENCE Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
CONTACT: Chris Town

## COMMENT

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

FEATURES Location/Qualifiers  
1..791

/organism="Brassica oleracea"  
/strain="T01000D3"  
/db\_xref="taxon:3712"  
/clone="B0GTC24"  
/clone\_lib="B0GT"

/note="Vector: pHS01, site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHS01 using BstXI linkers"

BASE COUNT 226 a 209 c 167 g 189 t  
ORIGIN

Query Match 79.1%; Score 17.4; DB 17; Length 791;  
Best Local Similarity 94.7%; Pred. No. 2.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 ATATCTACATCGAAGCA 22

Db 71 ATATCTACATCGAAGCA 89

## RESULT 5

B96258 203 bp DNA linear GSS 31-MAR-1998  
LOCUS B96258  
DEFINITION T23P5TRB TAMU Arabidopsis thaliana genomic clone T23P5, DNA sequence.

ACCESSION B96258  
VERSION B96258.1 GI:2998734  
KEYWORDS GSS

## SOURCE

ORGANISM

thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidops.

1 (bases 1 to 203)

REFERENCE Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Linher, K., Golden, K., Berry, K., Granger, D., Sun, E., Wilde, C., Adams, M.D. and Venter, J.C.

A BAC End Sequence Database for Identifying Minimal Overlaps in Arabidopsis Genomic Sequencing. Update 3  
Unpublished (1997)  
Other\_GSS: T23P5TRB

## COMMENT

JOURNAL

Contact: Steve Rounsley  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: rounsley@tigr.org

Seq primer: M13 Reverse

Class: BAC ends

High quality sequence stop: 203.

FEATURES Location/Qualifiers  
1..203

/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="T23P5"  
/clone\_lib="TAMU"  
/sex="hermaphrodite"  
/note="Vector: BelobACII; site\_1: HindIII; site\_2: HindIII  
; Produced by Rod Wing"

BASE COUNT 71 a 42 c 30 g 60 t  
ORIGIN

Query Match 78.2%; Score 17.2; DB 17; Length 203;  
Best Local Similarity 86.4%; Pred. No. 2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 CGTATCTACATCGAAGCA 22  
Db 61 CTAATCTACATCGAAGCA 82

## RESULT 6

CNS00V67/c 660 bp DNA linear GSS 28-JUN-1999  
LOCUS CNS00V67  
DEFINITION Arabidopsis thaliana genome survey sequence SP6 end of BAC T914 of TAMU library from strain Columbia of Arabidopsis thaliana, genomic survey sequence.

ACCESSION A0092245  
VERSION A0092245.1 GI:5293399  
KEYWORDS GSS

## SOURCE

ORGANISM

Arabidopsis thaliana.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 660)

REFERENCE Salanoubat, M., Choisme, N., Artiguenave, F., Brotier, P., Wincker, P.,



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:04:24 ; Search time 165.879 seconds  
(without alignments)  
2147.949 Million cell updates/sec

Title: US-09-477-082-30  
Perfect score: 22  
Sequence: 1 cgtatctacatcgaacga 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estda:\*  
2: em\_esthm:\*  
3: em\_estin:\*  
4: em\_estun:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_estl2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estcom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rnd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17.8	80.9	213	10	BB228117 BB228117
C 2	17.8	80.9	944	12	BF978743 BF978743
C 3	17.8	80.9	1970	17	AG097750 Pan trogl
4	17.4	79.1	17		BH582702 BCGTC24TR
5	17.2	78.2	203	17	B96258 T23P5TRB TA
6	17.2	78.2	660	17	CNS00VG7 AL092245 Arabidops

C 7	17.2	78.2	817	17	BH530096
C 8	17	77.3	587	13	BJ156899
C 9	16.8	76.4	136	13	BJ379738
C 10	16.8	76.4	204	10	BB310076
C 11	16.8	76.4	324	9	AL501502
C 12	16.8	76.4	451	9	AU037723
C 13	16.8	76.4	592	14	BQ874972
C 14	16.8	76.4	639	17	AQ311480
C 15	16.8	76.4	766	17	BH384117
C 16	16.8	76.4	769	17	BH504028
C 17	16.8	76.4	769	17	BH504028
C 18	16.4	74.5	162	17	BH069966
C 19	16.4	74.5	588	17	BH550594
C 20	16.4	74.5	687	17	AQ259507
C 21	16.4	74.5	902	17	BH166972
C 22	16.2	73.6	297	9	AV149405
C 23	16.2	73.6	360	17	AQ099918
C 24	16.2	73.6	469	17	PT022M1U
C 25	16.2	73.6	543	10	AM030345
C 26	16.2	73.6	585	17	AZ000131
C 27	16.2	73.6	612	17	AQ399904
C 28	16.2	73.6	671	13	BJ079939
C 29	16.2	73.6	675	17	AG075807
C 30	16.2	73.6	736	17	BH436611
C 31	16.2	73.6	736	10	AM349438
C 32	16.2	73.6	773	12	BG468967
C 33	16.2	73.6	788	17	BH582706
C 34	16.2	73.6	829	12	BG784865
C 35	16.2	73.6	844	12	BF294617
C 36	16.2	73.6	941	12	BG545504
C 37	16.2	73.6	1000	17	CNS060XC
C 38	16.2	73.6	1665	14	BM912231
C 39	16	72.7	444	17	BH634285
C 40	15.8	71.8	183	10	BE181317
C 41	15.8	71.8	235	10	BB305012
C 42	15.8	71.8	306	17	BH842158
C 43	15.8	71.8	329	9	AV110535
C 44	15.8	71.8	393	13	B1244929
C 45	15.8	71.8	414	17	A2162690

## ALIGNMENTS

RESULT 1  
BB228117/c  
LOCUS BB228117  
DEFINITION BB228117 RIKEN full-length enriched, 3 days neonate thymus Mus musculus cDNA clone A630008P18 3' similar to DB5189 Rattus norvegicus mRNA for Acyl-CoA synthetase, mRNA sequence.

ACCESSION BB228117  
VERSION BB228117.1  
KEYWORDS GI:8898762  
SOURCE house mouse.  
ORGANISM Mus musculus.

REFERENCE  
AUTHORS Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arikawa, T., Carrinholi, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirazane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawaji, K., Kikuchi, N., Kiyosawa, H., Kojima, T., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugihara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamana, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE RIKEN Mouse ESTs (Kono, H., et al.)  
JOURNAL Unpublished (2000)  
COMMENT Contact: Yoshihide Hayashizaki

OS	Homo sapiens.
XX	
XX	WO200218632-A2.
XX	
XX	07-MAR-2002.
XX	
XX	01-SEP-2001; 2001WO-EPI0074.
XX	
XX	01-SEP-2000; 2000DE-1043826.
PR	05-SEP-2000; 2000DE-1044543.
XX	
PA	(EPIC-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K, Guetig D;
DR	WPI; 2002-371829/40.
XX	
PT	Determining the degree of cytosine methylation in genomic DNA, useful
PT	for diagnosis and prognosis, comprises selective hybridization of
PT	amplicons from chemically treated DNA -
XX	
PS	Claim 12; 56pp + Sequence Listing; 56pp; German.
CC	
CC	This invention describes a novel method for determining the degree of
CC	methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC	genomic sample of DNA. The sample is treated chemically to convert
CC	cytosine (C) but not methylated C, to uracil, then part of the genomic
CC	DNA that contains the target C is amplified to form a labeled amplicon.
CC	The amplicon is hybridised to two classes, each with at least one
CC	member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC	and the degree of hybridisation to both classes is determined from the
CC	label on the amplicon. From the ratio of labels hybridised to the two
CC	classes of oligomers, the degree of methylation is calculated. The method
CC	is used: (i) for diagnosis and/or prognosis of side effects of
CC	therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC	of the central nervous, cardiovascular, gastrointestinal and respiratory
CC	systems etc., particularly by detecting mutations or single nucleotide
CC	polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC	types and for investigating cell differentiation. The method allows the
CC	methylation status of many C residues to be determined simultaneously.
CC	ABQ3410-ABQ3412, represent genomic DNA sequences used to illustrate the
CC	method for determining the degree of cytosine methylation described in
CC	the disclosure of the invention.
XX	
SO	Sequence 598 BP; 184 A; 48 C; 105 G; 261 T; 0 other;
	Query Match            73.6%;   Score 16.2; DB 24;   Length 598;
	Best Local Similarity   85.7%;   Pred. No. 1.3e+02;
	Matches     18; Conservative   0; Mismatches     3; Indels       0; Gaps     0.
OY	1 CGTATCTCATCGGAACG 21
DB	451 CGTATTTTAATTCGAACG 471
RESULT 15	
ABO38893/C	
ID	ABO38893 standard; DNA; 598 BP.
AC	ABO38893;
XX	
DT	12-JUL-2002 (first entry)
DE	
XX	Oligonucleotide for detecting cytosine methylation SEQ ID NO 25484.
XX	
KM	Human: cytosine methylation: 5'-CpG-3'; uracil: cytosine; diagnosis:
KM	drug; side effect: cancer; central nervous system; cardiovascular;
KM	gastrointestinal; respiratory system; single nucleotide polymorphism;
KM	SNP; cell differentiation; ds.

```

XX OS Homo sapiens.
XX
XX EN WO200218632-A2.
XX
XX PD 07-MAR-2002.
XX
XX PF 01-SEP-2001; 2001WO-EPI0074.
XX
XX PR 01-SEP-2000; 2000DE-1043826.
XX
XX PR 05-SEP-2000; 2000DE-1044543.
XX
XX PA (EPIG-) EPIGENOMICS AG.
XX
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX WP: 2002-371829/40.
XX
XX
XX PT Determining the degree of cytosine methylation in genomic DNA, useful
XX PT for diagnosis and prognosis, comprises selective hybridization of
XX PT amplicons from chemically treated DNA.
XX
XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX
XX CC This invention describes a novel method for determining the degree of
XX CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX CC genomic sample of DNA. The sample is treated chemically to convert
XX CC cytosine (C) but not methylated C, to uracil, then part of the genomic
XX CC DNA that contains the target C is amplified to form a labeled amplicon.
XX CC The amplicon is hybridised to two classes, each with at least one
XX CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
XX CC and the degree of hybridisation to both classes is determined from the
XX CC label on the amplicon. From the ratio of labels hybridised to the two
XX CC classes of oligomers, the degree of methylation is calculated. The method
XX CC is used: (i) for diagnosis and/or prognosis of side effects of
XX CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
XX CC of the central nervous, cardiovascular, gastrointestinal and respiratory
XX CC systems etc., particularly by detecting mutations or single nucleotide
XX CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
XX CC types and for investigating cell differentiation. The method allows the
XX CC methylation status of many C residues to be determined simultaneously.
XX CC AB013410-AB054121 represent genomic DNA sequences used to illustrate the
XX CC method for determining the degree of cytosine methylation described in
XX CC the disclosure of the invention.
XX
XX
XX SQ Sequence 598 BP; 261 A; 105 C; 48 G; 184 T; 0 other:
XX
XX
XX Query Match 73.6%; Score 16.2; DB 24; Length 598:
XX Best Local Similarity 85.7%; Freq. No. 1,3e+02;
XX Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0:
XX
XX
XX QY 1 CGTATATCTACATTGGAACG 21
XX ||||||| |||||||
XX DB 148 CGTATATTTTAATTCGAACG 128

```

XX 12-JUL-2002 (first entry)  
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 30313.  
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
XX drug; side effect; cancer; central nervous system; cardiovascular;  
XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
XX SNP; cell differentiation; ds.  
XX Homo sapiens.  
XX WO200218632-A2.  
XX 07-MAR-2002.  
XX 01-SEP-2001; 2001WO-EP10074.  
XX 01-SEP-2000; 2000DE-1043826.  
XX 05-SEP-2000; 2000DE-1044543.  
XX (EPIG-) EPIGENOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX WPI; 2002-371829/40.  
XX Determining the degree of cytosine methylation in genomic DNA, useful  
XX for diagnosis and prognosis, comprises selective hybridization of  
XX amplicons from chemically treated DNA -  
XX Claim 12: 56pp + Sequence Listing; 56pp; German.  
XX This invention describes a novel method for determining the degree of  
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
XX genomic sample of DNA. The sample is treated chemically to convert  
XX cytosine (C) but not methylated C, to uracil, then part of the genomic  
XX DNA that contains the target C is amplified to form a labeled amplicon.  
XX The amplicon is hybridised to two classes, each with at least one  
XX member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
XX and the degree of hybridisation to both classes is determined from the  
XX label on the amplicon. From the ratio of labels hybridised to the two  
XX classes of oligomers, the degree of methylation is calculated. The method  
XX is used: (i) for diagnosis and/or prognosis of side effects of  
XX therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
XX of the central nervous, cardiovascular, gastrointestinal and respiratory  
XX systems etc., particularly by detecting mutations or single nucleotide  
XX polymorphisms (SNP's) and (ii) for differentiation of cell or tissue  
XX types and for investigating cell differentiation. The method allows the  
XX methylation status of many C residues to be determined simultaneously.  
XX AB013410-AB054121 represent genomic DNA sequences used to illustrate the  
XX method for determining the degree of cytosine methylation described in  
XX the disclosure of the invention.  
XX Sequence 554 BP; 99 A; 57 C; 168 G; 230 T; 0 other;  
XX  
XX Query Match 73.6%; Score 16.2; DB 24; Length 554;  
XX Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
XX Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
XX  
XX 1 CGTATATCTACATTCGAAACG 21  
XX DB 109 CGTAAATATTAATTCGAAACG 89

RESULT 13  
ABQ43723  
ID ABQ43723 standard; DNA; 554 BP.  
AC ABQ43723;  
XX 12-JUL-2002 (first entry)  
XX

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 30314.  
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
XX drug; side effect; cancer; central nervous system; cardiovascular;  
XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
XX SNP; cell differentiation; ds.  
XX Homo sapiens.  
XX WO200218632-A2.  
XX 07-MAR-2002.  
XX 01-SEP-2001; 2001WO-EP10074.  
XX 01-SEP-2000; 2000DE-1043826.  
XX 05-SEP-2000; 2000DE-1044543.  
XX (EPIG-) EPIGENOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX WPI; 2002-371829/40.  
XX Determining the degree of cytosine methylation in genomic DNA, useful  
XX for diagnosis and prognosis, comprises selective hybridization of  
XX amplicons from chemically treated DNA -  
XX Claim 12: 56pp + Sequence Listing; 56pp; German.  
XX This invention describes a novel method for determining the degree of  
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
XX genomic sample of DNA. The sample is treated chemically to convert  
XX cytosine (C) but not methylated C, to uracil, then part of the genomic  
XX DNA that contains the target C is amplified to form a labeled amplicon.  
XX The amplicon is hybridised to two classes, each with at least one  
XX member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
XX and the degree of hybridisation to both classes is determined from the  
XX label on the amplicon. From the ratio of labels hybridised to the two  
XX classes of oligomers, the degree of methylation is calculated. The method  
XX is used: (i) for diagnosis and/or prognosis of side effects of  
XX therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
XX of the central nervous, cardiovascular, gastrointestinal and respiratory  
XX systems etc., particularly by detecting mutations or single nucleotide  
XX polymorphisms (SNP's) and (ii) for differentiation of cell or tissue  
XX types and for investigating cell differentiation. The method allows the  
XX methylation status of many C residues to be determined simultaneously.  
XX AB013410-AB054121 represent genomic DNA sequences used to illustrate the  
XX method for determining the degree of cytosine methylation described in  
XX the disclosure of the invention.  
XX Sequence 554 BP; 230 A; 168 C; 57 G; 99 T; 0 other;  
XX  
XX Query Match 73.6%; Score 16.2; DB 24; Length 554;  
XX Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
XX Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
XX  
XX 1 CGTATATCTACATTCGAAACG 21  
XX DB 446 CGTAAATATTAATTCGAAACG 466

RESULT 14  
ABQ38892  
ID ABQ38892 standard; DNA; 598 BP.  
AC ABQ38892;  
XX 12-JUL-2002 (first entry)  
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 25483.  
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
XX

RESULT 10  
ABQ48138/c  
ID ABQ48138 standard; DNA; 513 BP.  
XX  
XX ABQ48138;  
AC  
XX  
XX 12-JUL-2002 (first entry)  
DT  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 34729.  
XX  
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KM drug; side effect; cancer; central nervous system; cardiovascular;  
KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
KM SNP; cell differentiation; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO200218632-A2.  
PN  
XX 07-MAR-2002.  
PD  
XX 01-SEP-2001; 2001WO-EP10074.  
PF  
XX 01-SEP-2000; 2000DE-1043826.  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
XX (EPIC-) EPICENOMICS AG.  
PA  
XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
PI WPI; 2002-371829/40.  
DR  
XX  
XX Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA  
XX  
XX  
XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
PS  
XX  
XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
XX  
XX Sequence 513 BP; 115 A; 41 C; 167 G; 189 T; 1 other;  
SQ

Query Match 74.5%; Score 16.4; DB 24; Length 513;  
Best Local Similarity 94.4%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ATATCTACATCGAAGC 21  
IIIIIIIIIIIIIIIIIIII  
DB 443 ATATCTACATCGAAGC 426

ID ABQ48139 standard; DNA; 513 BP.  
XX  
XX ABQ48139;  
AC  
XX  
XX 12-JUL-2002 (first entry)  
DT  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 34730.  
XX  
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KM drug; side effect; cancer; central nervous system; cardiovascular;  
KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
KM SNP; cell differentiation; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO200218632-A2.  
PN  
XX 07-MAR-2002.  
PD  
XX 01-SEP-2001; 2001WO-EP10074.  
PF  
XX 01-SEP-2000; 2000DE-1043826.  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
XX (EPIC-) EPICENOMICS AG.  
PA  
XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
PI WPI; 2002-371829/40.  
DR  
XX  
XX Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA  
XX  
XX  
XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
PS  
XX  
XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
XX  
XX Sequence 513 BP; 189 A; 167 C; 41 G; 115 T; 1 other;  
SQ

Query Match 74.5%; Score 16.4; DB 24; Length 513;  
Best Local Similarity 94.4%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ATATCTACATCGAAGC 21  
IIIIIIIIIIIIIIIIIIII  
DB 71 ATATCTACATCGAAGC 88

RESULT 12  
ABQ43722/c  
ID ABQ43722 standard; DNA; 554 BP.  
XX  
XX ABQ43722;  
AC

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 3 TATATCTACATTCGAAACGA 22  
|||||  
Db 496 TATATCTACATTCGAAACGA 515

RESULT 8  
ABQ29578/C  
ID ABQ29578 standard; DNA; 1447 BP.  
XX  
AC ABQ29578;  
XX  
DT 12-JUL-2002 (first entry)  
XX

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 16159.  
XX  
KM Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KM drug; side effect; cancer; central nervous system; cardiovascular;  
KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
KM SNP; cell differentiation; ds.  
XX  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP10074.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
XX  
PA (EPIC-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
XX  
DR WPI: 2002-371829/40.  
XX  
XX

PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA -  
XX  
XX

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
XX

Sequence 1447 BP; 353 A; 134 C; 306 G; 654 T; 0 other;

Query Match 76.4%; Score 16.8; DB 24; Length 1447;  
Best Local Similarity 90.0%; Pred. No. 72;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 3 TATATCTACATTCGAAACGA 22

Db 329 TATATATAAATTCGAAACGA 310  
|||||

RESULT 9  
ABQ29579  
ID ABQ29579 standard; DNA; 1447 BP.  
XX  
AC ABQ29579;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 16170.  
XX  
XX

KM Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KM drug; side effect; cancer; central nervous system; cardiovascular;  
KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
KM SNP; cell differentiation; ds.  
XX  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP10074.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
XX  
PA (EPIC-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
XX  
DR WPI: 2002-371829/40.  
XX  
XX

PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA -  
XX  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
XX

Sequence 1447 BP; 654 A; 306 C; 134 G; 353 T; 0 other;

Query Match 76.4%; Score 16.8; DB 24; Length 1447;  
Best Local Similarity 90.0%; Pred. No. 72;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 3 TATATCTACATTCGAAACGA 22  
|||||  
Db 1119 TATATATAAATTCGAAACGA 1138

CC provided.  
 XX Sequence 25 BP; 12 A; 6 C; 0 G; 7 T; 0 other;  
 SQ

Query Match 76.4%; Score 16.8; DB 21; Length 25;  
 Best Local Similarity 90.0%; Pred. No. 53;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TATATCTACATCGAAGCA 22  
 |||||  
 Db 6 TATATCTACATCGAAGCA 25

RESULT 6  
 ABQ42138/c  
 ID ABQ42138 standard; DNA; 522 BP.  
 XX  
 AC ABQ42138;  
 XX  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 28729.  
 XX  
 KW Human; cytosine methylation; 5'-CPG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200218632-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EP10074.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX  
 DR WPI; 2002-371829/40.  
 XX  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA  
 XX  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CPG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 XX  
 SQ Sequence 522 BP; 64 A; 70 C; 244 G; 141 T; 3 other;

Query Match 76.4%; Score 16.8; DB 24; Length 522;  
 Best Local Similarity 90.0%; Pred. No. 67;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TATATCTACATCGAAGCA 22  
 |||||  
 Db 27 TATATCTACATCGAAGCA 8

RESULT 7  
 ABQ42139  
 ID ABQ42139 standard; DNA; 522 BP.  
 XX  
 AC ABQ42139;  
 XX  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 28730.  
 XX  
 KW Human; cytosine methylation; 5'-CPG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200218632-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EP10074.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX  
 DR WPI; 2002-371829/40.  
 XX  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA  
 XX  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CPG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 XX  
 SQ Sequence 522 BP; 141 A; 244 C; 70 G; 64 T; 3 other;

Query Match 76.4%; Score 16.8; DB 24; Length 522;  
 Best Local Similarity 90.0%; Pred. No. 67;







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:03:20 ; Search time 21.949 Seconds  
(without alignments)  
2257.229 Million cell updates/sec

Title: US-09-477-082-30

Perfect score: 22

Sequence: 1 cgtatctcattcgaacga 22

Scoring table:

IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq\_101002:\*

- 1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*
- 2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*
- 3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*
- 4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*
- 5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*
- 6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*
- 7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*
- 8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*
- 9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*
- 10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*
- 11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*
- 12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*
- 13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*
- 14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*
- 15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*
- 16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*
- 17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*
- 18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*
- 19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*
- 20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*
- 21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*
- 22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*
- 23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*
- 24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	22	21	AA51819
2	17.8	80.9	11198	23	ABL10532
3	17.2	78.2	9289	22	AA546501
4	17.2	78.2	22191	20	AA220600
5	16.8	76.4	25	21	AA51821
6	16.8	76.4	522	24	ABQ42138
7	16.8	76.4	522	24	ABQ42139
8	16.8	76.4	1447	24	ABQ29578
9	16.8	76.4	1447	24	ABQ29579

C	10	16.4	74.5	513	24	ABQ48138	Oligonucleotide fo
C	11	16.4	74.5	513	24	ABQ48139	Oligonucleotide fo
C	12	16.2	73.6	554	24	ABQ43722	Oligonucleotide fo
C	13	16.2	73.6	554	24	ABQ43723	Oligonucleotide fo
C	14	16.2	73.6	598	24	ABQ38892	Oligonucleotide fo
C	15	16.2	73.6	598	24	ABQ38893	Oligonucleotide fo
C	16	16.2	73.6	719	24	ABQ27536	Oligonucleotide fo
C	17	16.2	73.6	719	24	ABQ27537	Oligonucleotide fo
C	18	16.2	73.6	2621	23	ABL26612	Drosophila melanog
C	19	16.2	73.6	6337	22	AA546450	Tumour suppressor
C	20	16.2	73.6	6337	22	ABL33311	Human immune syste
C	21	16.2	73.6	8979	23	ABL13054	Drosophila melanog
C	22	16.2	73.6	11295	23	ABL13064	Drosophila melanog
C	23	16.2	73.6	14902	23	ABL26614	Drosophila melanog
C	24	16.2	73.6	39651	23	ABL18856	Drosophila melanog
C	25	16	72.7	14245	23	ABL03938	Drosophila melanog
C	26	16	72.7	15238	23	ABL19570	Drosophila melanog
C	27	16	72.7	15243	23	ABL19568	Drosophila melanog
C	28	15.8	71.8	2007	24	ABQ67849	Listeria innocua D
C	29	15.8	71.8	2007	24	ABQ69467	Listeria innocua D
C	30	15.8	71.8	3123	23	ABL29094	Drosophila melanog
C	31	15.8	71.8	3399	16	AAQ91095	Human herpesvirus-
C	32	15.8	71.8	3777	23	ABL29200	Drosophila melanog
C	33	15.8	71.8	6849	24	ABL92253	Chemically treated
C	34	15.8	71.8	16275	23	ABL09786	Drosophila melanog
C	35	15.8	71.8	31449	23	ABL09172	Drosophila melanog
C	36	15.8	71.8	1163020	24	ABQ67197	Listeria innocua c
C	37	15.8	71.8	3011208	24	ABQ69245	Listeria innocua D
C	38	15.6	70.9	384	22	AA187653	Human polynucleoti
C	39	15.6	70.9	526	24	ABQ19854	Oligonucleotide fo
C	40	15.6	70.9	526	24	ABQ19855	Oligonucleotide fo
C	41	15.6	70.9	535	24	ABQ48358	Oligonucleotide fo
C	42	15.6	70.9	535	24	ABQ48359	Oligonucleotide fo
C	43	15.6	70.9	632	24	ABQ37584	Oligonucleotide fo
C	44	15.6	70.9	632	24	ABQ37585	Oligonucleotide fo
C	45	15.6	70.9	1221	24	ABQ41560	Oligonucleotide fo

# ALIGNMENTS

AA51819	standard; DNA; 22 BP.
AA51819:	
AC	AA51819:
XX	
DT	31-OCT-2000 (first entry)
XX	
DE	Antisense primer for Casp8 methylated-specific disulfite treated DNA.
XX	
KW	CASP8; caspase-8; promoter; inactivation; methylation; cysteine protease;
KW	tumour suppressor; chromosome 2q33-34; neuroblastoma; cancer;
KW	death receptor; apoptosis; cytosolic; gene therapy; primer; ss.
XX	
OS	Homo sapiens.
XX	
PM	WO200039347-A1.
XX	
PD	06-JUL-2000.
XX	
PF	30-DEC-1999; 99WO-US31280.
XX	
PK	31-DEC-1998; 98US-0114308.
XX	
PA	(SUSD-1) ST JUDE CHILDREN'S RES HOSPITAL.
XX	
PI	Kidd VJ, Lahti JM, Teitz T;
XX	
DR	WPI: 2000-452423/39.
XX	
PT	Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or
PT	prognosing cancer, comprises detecting a modification of genomic DNA

```

/product="hypothetical protein"
/protein_id="AAD23041.1"
/db_xref="gi:4559381"
/translation="MAESSTIARLYRCPCCKNLQEPKDCPPFOCGGCGTYLCAKN
DRADLVSKVENRSKELEVTSSEESDSNASLRHOQSGTEFWVDVHNDPSKMN
VDSGNDSQPLDQIRKRTKRCLEGRFSTSNFTDLSLDEAIEQDRAGLVKRL
DKLEQLOVPKEQOIPSSSSGFKAPLRYFSSGRYAGPSYQYSEPEPCLYNNNS
DVLHGPMHNPAVPAVGDHGFPMHEPTLQPSHLNSROYIGNNGHVDFTDHPHNGM
FHQSTCSWLPDALRNPGEYPHERSFEGFTSLSPRTFPPGSSPGPOLRGRTSGFK
DMRMNAVSRYHPKPYSSSGSRLLQPVAGAPFNCINCEKLIKPEKIDSATRKQ
RMRCGACSVLDYSFYDKLITSTDPASARKPETHRLRWETTANPSSDDYDNAIEF
HADRGSADVSTALVLISNTAQMQLTDAISPSVSEDELSDSSTVRAKYNLASPLAKQ
FEVSSINVRDRSGPRSQSSRSQDRTYLSKTAQONSMKASVAIEMDVNDYSHNNEV
SODSANDCTDDQGRTKKGFASIMKNSFKDLKSIONEGSDYSINGHFAEDLYKMA
EKQAGPIRPGNMYDYRAGFWGLGSHCLGILPPFIEELNYPMPENCAGGTRFVNG

```

Query Match 78.2%; Score 17.2; DB 8; Length 95214;  
 Best Local Similarity 86.4%; Pred. No. 5.4e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGTATATCTACATTCGAAACGA 22  
 | ||||| |||| |||||  
 Db 82489 CTTATATCTGCATTGAAACGA 82510

Search completed: July 8, 2003, 04:02:41  
 Job time : 83.3011 secs



TAFTIGTICMRETTLSKSLVEBMEVDAQEAVLPDGTSEAFLEALSDDVNDYHLHDSQEH  
 " complement(14394. .15192)  
 /gene="At2g46505"  
 /note="F13A10.20; supported by full length cDNA: Ceres:  
 11006"  
 mRNA complement(<14394. .>15192)  
 /gene="At2g46505"  
 complement(14656. .15111)  
 /gene="At2g46505"  
 /codon\_start=1  
 /product="Expressed protein"  
 /protein\_id="AA15243.1"  
 /db\_xref="GI:2019780"  
 translation="MSLRRTLDLHROTORATLSKSPSMHSSASATAVRNPLG  
 RLSSITPFAOAKLKPDSTNYLDRSSISSGTSPLNKAARSSRSSRGYTNCSFLR  
 KIPVPHIHESMEILLADYHDEMRNLVMSLGFQIIVLKDITLPLT"  
 complement(15830. .17530)  
 gene /gene="At2g46510"  
 /note="F13A10.4"  
 mRNA complement(<15830. .>17530)  
 /gene="At2g46510"  
 complement(15830. .17530)  
 /gene="At2g46510"  
 /codon\_start=1  
 /product="putative bHLH transcription factor"  
 /protein\_id="AA20162.2"  
 /db\_xref="GI:2019775"  
 translation="NMNMDLGDWDEDDKSVSAVGLHSLASDFLRANSNONTFLMGT  
 DDTLNKSLSLVDWNSSEFNNRYALFMQOTISREGQVYLGSGDCCEPNEEESKY  
 VRSYNNNGASEETWQDMKRVYLOKHLHLEGGSDNDALSLKVATETFLFSLMS  
 PFNHEGGGRGRCYSSGKHVMSLDVANSDDCFRFSMAKSAGIRTVAVPDAVLE  
 LGSVMSLPIENIGLYKSVQALFMRVYIQPMVTSNNMGIHKLFGODLSGAHAPK  
 LEVRNLDERETPOSMEGYNKNGKPTFGTTPORDVYVDENNVMDNNNYKTOLEPA  
 GSVVAASNPSTNTOEESCESEKRPVSLAGAGVSVDEKRPKRKPAENRE  
 PLNHVABORREKLNORRYALRSVYPNISKMDKSLSDATSYIKELQEKXKIMDE  
 RGTGDKSLESNTIYEESEPVDIQAMEEYVRYISPLDSPARITIQAMNSVSL  
 MEKLSLADDTMHTFVILKSNMGSDPLTEKLLAAYPTSTQPPPLSSQSVGDI  
 " repeat\_region  
 17920. .17946  
 /rpt\_family="(GA)n"  
 repeat\_region complement(18392. .18454)  
 /rpt\_family="AT\_rich"  
 gene 21510. .24428  
 /gene="At2g46520"  
 /note="F13A10.5"  
 mRNA <21510. .>24428  
 /gene="At2g46520"  
 21510. .24428  
 /gene="At2g46520"  
 /codon\_start=1  
 /product="putative cellular apoptosis susceptibility  
 protein"  
 /protein\_id="AA20163.1"  
 /db\_xref="GI:4415933"  
 translation="MEMRETLVETLSOCFLNTLSPPEPRTAERALSADAQANGL  
 AVRIYAFAPIDQCTRHAAVNNKHLBRNWRHAGSGSPYDSEKEDIKITVISM  
 LSAAPRISOLSEALVYTGKHPKAMPALPEPLANLQNALAGVSVNLTIGAS  
 SIKRFSYERIDALFVLDKLDNFAPALTEFLKTSLLISAASSGSPILPLF  
 ESORLCITFYSLFQDLPEFEDHMKEMGKTKLSNYPALSETEGTLVLDLR  
 AALCENINNIENKEEEOGFLNEFASVMTLIDVSKSPSDOGLTAKLTST  
 SVHHLFAGDNVKEICOSIVPNGLRDEDEIFEMNYIEFRRMESDQVLEST  
 ACELLGATNRYKTYEYVSLERLISFSFANPSAMKDKDCAIYIVYSTKAC  
 GASVSTDLIDVQNFANITLPELQSDVNSPMLKAGSLKFLTMESHLPKPAQVL  
 PELVRLKESNVSHVSAECLEKLLVEEAGRNRYAGDLPPLDLMNEDAL  
 KPESEENYIANKIMRYGVADISAENVAGPCIGGLTSLSEVCKNKPINNHLE  
 SVAVLRACERDILISAFETSLPSLOMILANDITEPLPYGQQLADVLNPLT  
 SPYVQIFLLISPEMSKRGVPAVRLLOAFLOKAPHEVOENKSLQVIGLEKLV  
 ASPSDQEGEYILNTIENLDSVIAVPMKGVMSALFTVQNKTKVKKQKSVIFKSL  
 FLVKGCAVIVETMTVQPNITIAIVHEWMIRKLKIMSMYKFLAVATILCTEP  
 ALIDPSAKLMGKMLDSIVTLVSRPEQERVDPEPELISENGITAAVAKLHNGKK  
 EEDPLKIDPKQFLVASVSRSSASPGRYPTQIGENLEQANOTALIQLCNAYNGCIA  
 "

repeat\_region 25088. .25183  
 /rpt\_family="Rf:g11100803|emb|X93607.1|AT74A A.thaliana  
 microsatellite repeat region (clone 74A)#Rf:ATR0046|X93607  
 74A"  
 repeat\_region 25090. .25183  
 /rpt\_family="(GA)n"  
 repeat\_region 25090. .25166  
 /rpt\_family="Rf:g11100803|emb|X93607.1|AT74A A.thaliana  
 microsatellite repeat region (clone 74A)#Rf:ATR0046|X93607  
 74A"  
 repeat\_region 25090. .25162  
 /rpt\_family="Rf:g11100803|emb|X93607.1|AT74A A.thaliana  
 microsatellite repeat region (clone 74A)#Rf:ATR0046|X93607  
 74A"  
 repeat\_region 25090. .25158  
 /rpt\_family="Rf:g11100803|emb|X93607.1|AT74A A.thaliana  
 microsatellite repeat region (clone 74A)#Rf:ATR0046|X93607  
 74A"  
 repeat\_region 25090. .25154  
 /rpt\_family="Rf:g11100803|emb|X93607.1|AT74A A.thaliana  
 microsatellite repeat region (clone 74A)#Rf:ATR0046|X93607  
 74A"

Query Match 78.2%; Score 17.2; DB 8; Length 84825;  
 Best Local Similarity 86.4%; Pred. No. 5.4e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGTATCTCATTCGAAACGA 22  
 Db 23427 CTTATCTCGATTGAAACGA 23448

RESULT 15  
 AC006526  
 LOCUS  
 DEFINITION Arabidopsis thaliana chromosome 2 clone F1C10 map C1C02E07,  
 complete sequence.  
 ACCESSION AC006526  
 VERSION AC006526.8 GI:20197817  
 KEYWORDS HTG.  
 SOURCE Arabidopsis thaliana.  
 ORGANISM Arabidopsis thaliana.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 1 (bases 1 to 95214)  
 Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,  
 Barnstead,M.E., Mason,T.M., Bowman,C.L., Renning,C.M.,  
 Benito,M.-I., Carrera,A.J., Creasy,T.H., Bell,C.R., Town,C.D.,  
 Nierman,W.C., Fraser,C.M. and Venter,J.C.  
 Unpublished  
 2 (bases 1 to 95214)  
 Lin,X.  
 Direct Submission  
 Submitted (09-MAR-2000) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA  
 3 (bases 1 to 95214)  
 Town,C.D. and Kaul,S.  
 Direct Submission  
 Submitted (27-FEB-2002) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA, cdtown@igf.org

## COMMENT

On Oct 6, 1997 this sequence version replaced gi:2443866.  
 We have determined that YAC YUP812 is chimeric, and is comprised  
 of two distinct genomic EcoRI fragments from chromosome 1. This  
 submission contains the sequence from the EcoRI site at position 1  
 (right end) to position 181918 of our previous Phase II  
 submission YUP812 accession number AC000098. This fragment maps  
 at the bottom of the right arm of the chromosome between the ATHTAPS  
 and m532 markers.  
 Location/Qualifiers

## FEATURES

[illegible]

## JOURNAL

## COMMENT

Submitted (19-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org  
On Jan 19, 2001 this sequence version replaced g1:12280860.  
Address all correspondence to: atetigr.org

BAC clone T32E8 is from Arabidopsis thaliana chromosome 1  
The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, <http://CCR-081.mt.edu/GENSCAN.html>), GenemarkHM (Mark Borodovsky, <http://genemark.biology.gatech.edu/genemark/>), GlimmerA (a variant of GlimmerM, see Mihaela Pertea, <http://www.tigr.org/softlab/glimmer.htm>), and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact [mpertea@tigr.org](mailto:mpertea@tigr.org)), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

## FEATURES

## Source

1. 82454  
/organism="Arabidopsis thaliana"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="1"  
/clone="T32E8"  
complement(1..3320)  
/gene="T32E8.1"  
/note="similar to ribonuclease II RNB family protein GB:CA16367 [Schistosoma mansoni]; Pfam HMM hit: VacB and RNase II family 3'-5' exonuclease"  
complement(join(<1..69,154..2280,2381..2449,2524..2577,2676..3055,3215..3320))  
/gene="T32E8.1"  
/codon\_start=1  
/product="putative 3'-5' exonuclease, 3' partial; 3320-1"  
/protein\_id="AA051632.1"  
/db\_xref="GI:12323300"  
/translation="MKSASDSQSYERLENGHKKRRNRPQKQNRKSSQSVIEDAHE ESLDRDSRSKADKSTSSKQSPNDLEPAMRASNAFMSMPMAESQYPPRRAS PLSPSEVSKQLSKSCDPDPRACQSPNGELFQOIGSSQKIFESHWGLDVTAL LEKGAFKALFRVNAHNEAYCKIDGVPDILINGNCSRAVEGDVVKIDPLSL PMKMGVETSAARCEGNSPPEKDDKAROKNIDVVEGPDGFSKNSVITKRGK NGVTPSSPSIDSLGSCFCEKQKNSANDKICGLISFPKRPFGQVAYVEKLVND SIVGLDVGWVHTKESDPKCKSPLSLSDVEYVQLMPADRFPKLLVPHVLEGSTA ARLENDPLNEALEVAAQIVDWGSGSPVQAITHLFGSGSELPQNALIYQNSVD SDFSGSLTSLPVPWEVPEEVEVQRKLRDLQVLTIDPSTATOLDALSVQSLPGF FRGVGHADVSFVLPETALDEAFRSTSYLWQKISMPLPSEVNGSLGSGAR LAESILMDLNREGVIDRMIGRTIIRSCCKLSYHAODIDGKSDVAENGPALHGSF KMDQVTRSVKQSLSTILROKRRFNALQENKRPVLPDEHGVDPYDVPVCSKQNS FLVEEFLLANMTAAEVISQAYRASSLIRRPETRTKLKFEFGSCSHGDDLISSS GOLDSLEKLTGNLKDSDVEVDLNNVAKIRQMOLASIFCTGNLKDVAEWCHVALAP LYHFTSPLRPDIIVVRAALAALEAEELISKQKQTAIDGRSCFPGIIFHNKAABE IEKGKALSVAAALKGVGPSTETILSDVAAACNENKRLAARVADACDKLYTWFLVKKEIF PCFARVNMIGSRFMTVYISKI"  
1..189  
/note="overlap with BAC clone T5M16  
[AC010704:103165..103353].  
complement(18589..10716)  
/gene="T32E8.2"

/note="similar to AUX1 (regulator of root gravitropism, putative permease) GB:CA67308 [Arabidopsis thaliana]"  
complement(join(<8589..8783,9000..9465,9555..9652,9753..9936,10020..10132,10336..10421,10504..>10716))  
/gene="T32E8.2"  
complement(join(8589..8783,9000..9465,9555..9652,9753..9936,10020..10132,10336..10421,10504..10674))  
/gene="T32E8.2"  
/codon\_start=1  
/product="putative AUX1-like permease; 10674-8589"  
/protein\_id="AA051630.1"  
/db\_xref="GI:12323298"  
/translation="MAAEKIEIVVAGNYEMEREENISGNKSSYTKLSNFEWGG SYVDAMFSCSNVOAVLLELPFSGLGMSGLPOLFGYGLMSWYATLISLVLEY RTKREKEDFERHNVHOMFEVDGLGKHNRLNGLFGTGLFSGVLIOLCASNIY YINDKIDKRWYIRGACATYFTFTSPNLSMTGLAMTYLMTWYILSLHGO AEVKSAGPTWYLTITGATNLIYTGNAHYEINHAMKQKREKATILATLYVL LTPSASAYWAGDKDLTHSNLSLSPTEGFDVAILMLHQRTSGFASPTLYLA MERLIVHEKSNFKNRMAKRLPVVYIMFLAIIFPFGGINSVGLSVFTVYIIPA LAHMLFAPAPSNVNERPPRVYGGMGTYCINIFVYVWVFVGGFSGMASWNEF RQIDTFLGFKCYCCPKP"  
12104..13574  
/gene="T32E8.3"  
/note="similar to thaumatin-like protein GB:AA071214 [Arabidopsis thaliana]; Pfam HMM hit: Thaumatin family"  
join(<12104..12179,12505..12628,12704..>13574)  
/gene="T32E8.3"  
join(12104..12179,12505..12628,12704..13574)  
/gene="T32E8.3"  
/codon\_start=1  
/product="thaumatin-like protein; 12104-13574"  
/protein\_id="AA051631.1"  
/db\_xref="GI:12323299"  
/translation="MDRTIEFDINGPYDLSLRPNISVASLRSNNITFDPRIFS SKDANSREYTKDLVLSHILRWJTIIRPVYSGTKMSERATLINSQDQIM PATTPGNEGGGFEELKPGQSYFHAHPGSGVSGTGCSDSTGTCGSGOST LKCSASGKPPASLAETFLAHLDDYDVLSDGFEPLPSVPMNKSGCSACGYADLRP HCPQELAVKSNKGVICRSACDVFDEYCCRCVGNPVVCCPTYSKIFKQCPYAY SVYDPTSLMTCITADYVLSFCSRFVLEFTLTSHDYFENDSNQVSLDHDYIY MIYFIHPLISYAKVVIQWQK"  
14054..15303  
/gene="T32E8.4"  
/note="similar to hypothetical protein GB:p34661 [Caenorhabditis elegans]"  
join(14054..14174,14251..14348,14779..14811,15170..15303)  
/gene="T32E8.4"  
join(14107..14174,14251..14348,14779..14811,15170..15252)  
/gene="T32E8.4"  
/codon\_start=1  
/product="unknown protein; 14107-15252"  
/protein\_id="AA051633.1"  
/db\_xref="GI:12323301"  
/translation="MATGKRSKRVLTSDPKLPFVFSVPEGAPPTAVLKFAAEDEFK VPQTSALITNDIGINPQOSAGNVFLKHGSELRLIPRDVGAVFVMDP"  
18583..21388  
/gene="T32E8.5"  
/note="Pfam HMM hit: Eukaryotic protein kinase domain, C-terminal region is identical to putative protein Kinase GB:CA042680 [Arabidopsis thaliana]"  
join(<18583..18916,19056..19948,20068..20268,20382..20543,20662..21234)  
/gene="T32E8.5"  
join(18583..18916,19056..19948,20068..20268,20382..20543,20662..21234)  
/codon\_start=1  
/product="putative protein kinase; 18583-21234"  
/protein\_id="AA051619.1"  
/db\_xref="GI:12323287"  
/translation="MOTRSGPRSVAPQROASRNRIYTAEGORSADYVTTSSQSLA NTLQDTNNMLAITSVAGESASITQPTVSESRNPSPRDQFGASNTSLESNDYRK SOSTLIGTSQDEWDATQAEASHDQAGISGHQNLVSDEVSLEKSEYKQSSSLAKIQ GOLGEFPNFINOPRTCSAVGSSWATTTLLHSSAPMLNATTVHSRSYVEADSNAPH



/product="ribosomal protein S18 (rpsr)"  
 /protein\_id="AAC65056.1"  
 /db\_xref="GI:3322319"  
 /translation="MAEDHPSVDLTHISSPSESESAKKNROFRRKKVCEPCQKL  
 LADYKDPDLIRRTTERGKILPRITGTCARHQRVALEKRSRAVALLPVLITE"  
 complement(2593. .3123)  
 /gene="TP0062"  
 /note="TP0062"  
 /gene="TP0062"  
 /note="TP0062"  
 /note="similar to GB:AE000783 percent identity: 48.15;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="single-strand DNA binding protein (ssb)"  
 /protein\_id="AAC65057.1"  
 /db\_xref="GI:3322320"  
 /translation="MADVNRYVVGRLTDALDKTYSAGALCFESVAINRRKSGD  
 WVEVNEFDIVLMRGQEVTSYLTKGQVAVGELGQSRWEGQSRKVELSRINV  
 QILGSYVGARADEGFESSRYAAESTAGRVGTSDDSRPTGILGKRGLDARTS  
 LDADFSDDLDPVP"  
 complement(3132. .3413)  
 /gene="TP0063"  
 /note="TP0063"  
 /note="complement(3132. .3413)  
 /gene="TP0063"  
 /note="similar to GB:AE000783 percent identity: 30.00;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ribosomal protein S6 (rpsf)"  
 /protein\_id="AAC65058.1"  
 /db\_xref="GI:3322321"  
 /translation="MRTYELMAVSAHEDLFIQSGAVRALLQENDAVIAREDHIGER  
 ELAVPLKQKRGRLFTVQCEPKVKEHDKRLRLHDLTHLFRVDS"  
 3555. .4130  
 /gene="TP0064"  
 /note="TP0064"  
 /gene="TP0064"  
 /note="hypothetical protein; identified by Glimmer;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="T. pallidum predicted coding region TP0064"  
 /protein\_id="AAC65064.1"  
 /db\_xref="GI:3322327"  
 /translation="MLPTAQRRLRGYTRACAPYGRGEEPRRRCALCARVGHVCG  
 RSVQNFQPDNIYHARSIDVLEGQLMDCLYAEKVYDVEFEDHALQCLAH  
 DRGHPDYLPLICLFCGCAFLNLNLFHRAAPVTAQVEFTMLPFEKRHSALH  
 ENLARSIOEVDTSADASHVVSODEIDELLEH"  
 4204. .4791  
 /gene="TP0065"  
 /note="TP0065"  
 /note="similar to GP:1694865 percent identity: 36.11;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAC65059.1"  
 /db\_xref="GI:3322322"  
 /translation="MLKNHYLRCPDGPIRPAMDRESLEFALLGDMGSCFLDLFAS  
 GVGLEAVSRGAVPVFVEMVRSFVLLQVVALGRCRCVAVRYTARATLTH  
 FVYLDPPEPVFHAELLQRTSRASLCREGSVVMVHRERREKKLADKIDSLVRTQRYVG  
 RSVDFTRIKPVLRRTETRCILLFVRGDAEGA"  
 4775. .5089  
 /gene="TP0066"  
 /note="TP0066"  
 /note="TP0066"  
 /note="hypothetical protein; identified by Glimmer;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="T. pallidum predicted coding region TP0066"  
 /protein\_id="AAC65065.1"

/db\_xref="GI:3322328"  
 /translation="MQRVLESDFYFVNGIGRPVSTSDRDRALLNRGNAYLNEGL  
 CEARVFTTGXDGLTRIGDYMKADVLTALRFYFAENQKMRPIVSALVILIC  
 LI"  
 5077. .6258  
 /gene="TP0067"  
 /note="TP0067"  
 /note="similar to GB:AE000783 percent identity: 47.91;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAC65060.1"  
 /db\_xref="GI:3322323"  
 /translation="MSYIILTPGGGMLGCECVPPFETVETGKRSVDADRIMLSR  
 DGAFLKYNDEARASEGKILQIADNNYALVAGDAARARAYOASDYTRCLQ  
 YRNSYALFGLDNCYKNNRBYKAYEVMCOYEDSHNIVATPRADAYRKIDHONS  
 RUYISOVLNDEHNSYALIGLAHLIDYDKKIREALITYKMLLECAHSDVDIRLITIG  
 NCYRKMKLESQGLPYFQELAKRDGNEFYGFQMAQCIYGMKNQHSIOYERKILEKOT  
 ONRVILITRADAYRAHIGEEKAHQYKRALDIDYAVATGLAVLCKLQGREAVVS  
 LDRLVQLDKNNRYVVELADCYRKLQKOKALETLRPPQFGVKNRYVSELNSELEGA  
 S"  
 6439. .7461  
 /gene="TP0068"  
 /note="TP0068"  
 /note="similar to PID:1001734 PID:1208474 percent  
 identity: 39.39; identified by sequence similarity;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAC65061.1"  
 /db\_xref="GI:3322324"  
 /translation="MEMCCALSGLLPEIQRKVCAPAFERRRGVQVFRMIAAGCTDEHAM

Query Match 78.2%; Score 17.2; DB 1; Length 11101;  
 Best Local Similarity 86.4%; Pred. No. 5.7e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	1	CGATATCTACATTCGAACGA	22
Db	1		4466
		CATACATCATTCGAACGA	

RESULT 13  
 AC012193  
 LOCUS  
 DEFINITION  
 Arabidopsis thaliana chromosome 1 BAC T32E8 genomic sequence,  
 complete sequence.  
 AC012193  
 VERSION  
 AC012193.6  
 GI:12323286  
 KEYWORDS  
 HTG.  
 SOURCE  
 Arabidopsis thaliana.  
 ORGANISM  
 Arabidopsis thaliana.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 82454)  
 REFERENCE  
 Lin,X., Kaul,S., Town,C.D., Beutlo,M.-I., Greasy,T.H., Haas,B.J.,  
 Wu,D., Maitl,R., Rohnig,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,  
 Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.  
 Arabidopsis thaliana chromosome 1 BAC T32E8 genomic sequence  
 Unpublished  
 TITLE  
 JOURNAL  
 2 (bases 1 to 82454)  
 REFERENCE  
 Lin,X. and Kaul,S.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (21-OCT-1999) The Institute for Genomic Research, 9712  
 Medical Center Dr. Rockville, MD 20850, USA, xlnet@ig.org  
 3 (bases 1 to 82454)  
 REFERENCE  
 Town,C.D. and Kaul,S.  
 TITLE  
 Direct Submission

KPESEENQYLNKICIMRVLGADISAEVAGPCIGLTSLISEVCKNPNPIFNHYLE  
 SVAIVLRACEDISLISAFETSLPESLOMIANITETLEPQOLLOVEINPTL  
 SPNVMQIELLISPEWSKSGNVPALVRLQAFLOKAPREYQENRLOVIGIEPKIV  
 ASPSDDEGCFYIINTIENLDSVIAIPYKQMSALFPRVQKRYKQKQSVITMSL  
 FLVHQAIVLETNIVQPNITIAIVEHWIENLILMSSEVKILIAVAATRLICEIP  
 ALLDPSAAALMGKMDISIVTLVSREQEVLDPEMPELSENVGTAFAVKILHNGK  
 EDDPLDIDIPKQFIVASVRSASPGRYPIIIGENEOANOTLIOLCNAYNGCIA  
 "

BASE COUNT 753 a 618 c 689 g 859 t  
 ORIGIN  
 Query Match 78.2%; Score 17.2; DB 8; Length 2919;  
 Best Local Similarity 86.4%; Pred. No. 6e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 CGTATCTACATCGAAGA 22  
 Db 1918 CTTATCTGCAATTGAAGA 1939

RESULT 11  
 AX251255/c AX251255 9289 bp DNA linear PAT 05-OCT-2001  
 LOCUS Sequence 223 from Patent WO0168912.  
 ACCESSION AX251255  
 VERSION AX251255.1 GI:15984678  
 KEYWORDS  
 SOURCE .  
 ORGANISM synthetic construct.  
 REFERENCE 1 (bases 1 to 9289)  
 AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.  
 TITLE Diagnosis of diseases associated with tumor suppressor genes and  
 oncogenes  
 JOURNAL Patent: WO 0168912-A 223 20-SEP-2001;  
 Epigenomics AG (DE)  
 FEATURES  
 source Location/Qualifiers  
 1..9289  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="chemically treated genomic DNA (Homo sapiens)"  
 BASE COUNT 2515 a 206 c 2097 g 4471 t  
 ORIGIN  
 Query Match 78.2%; Score 17.2; DB 6; Length 9289;  
 Best Local Similarity 86.4%; Pred. No. 5.8e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 CGTATCTACATCGAAGA 22  
 Db 5667 CATATCTACATCCGATCGA 5646

RESULT 12  
 AE001191/c AE001191 11101 bp DNA linear BCF 16-JUL-1998  
 LOCUS Treponema pallidum section 7 of 87 of the complete genome.  
 ACCESSION AE001191 AE000520  
 VERSION AE001191.1 GI:3322316  
 KEYWORDS  
 SOURCE .  
 ORGANISM Treponema pallidum.  
 REFERENCE 1 (bases 1 to 11101)  
 AUTHORS Fraser, C.M., Norris, S.J., Weinstock, G.M., White, O., Sutton, G.G.,  
 Dodson, R., Gwin, M., Hickey, E.K., Clayton, R., Ketchum, K.A.,  
 Sodergren, E., Hardham, J.M., McLeod, M.P., Salzberg, S., Peterson, J.,  
 Khalak, H., Richardson, D., Howell, J.K., Chidambaram, M.,  
 Uterback, T., McDonald, L., Artlich, P., Bowman, C., Cotton, M.D.,  
 Venter, J.C. et al  
 TITLE Complete genome sequence of Treponema pallidum, the syphilis  
 spirochete  
 JOURNAL Science 281 (5375), 375-388 (1998)

MEDLINE 98332770  
 PUBMED 9665876  
 REFERENCE 2 (bases 1 to 11101)  
 AUTHORS Fraser, C.M., Norris, S.J., Weinstock, G.M., White, O., Sutton, G.G.,  
 Dodson, R., Gwin, M., Hickey, E.K., Clayton, R., Ketchum, K.A.,  
 Sodergren, E., Hardham, J.M., McLeod, M.P., Salzberg, S., Peterson, J.,  
 Khalak, H., Richardson, D., Howell, J.K., Chidambaram, M.,  
 Uterback, T., McDonald, L., Artlich, P., Bowman, C., Cotton, M.D.,  
 Fujii, C., Garland, S., Hatch, B., Horst, K., Roberts, K., Matthey, L.,  
 Weidman, J., Smith, H.O. and Venter, J.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-MAR-1998) The Institute for Genomic Research, 9712  
 Medical Center Dr, Rockville, MD 20850, USA  
 FEATURES  
 source Location/Qualifiers  
 1..11101  
 /organism="Treponema pallidum"  
 /db\_xref="taxon:160"  
 /complement(68..1384)  
 /gene="TP0058"  
 /complement(68..1384)  
 /gene="TP0058"  
 /complement(68..1384)  
 /note="similar to GB:D26185.SP:P37469.PID:467330  
 GB:AL009126 percent identity: 39.68; identified by  
 sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=1  
 /product="replicative DNA helicase (dnab)"  
 /protein\_id="AAC65054.1"  
 /db\_xref="GI:3322317"  
 /translation="MFGKPNFTQELKGIPEPHNEAEKRVAGVAVLDDSLATTELD  
 SASSTYSAAHQRIQDALVESLDGQRPDITLSLHSCENLDFVGSATVASTIDAV  
 PSAANVEYVTRIVCDAMRSLKLVARIIEAENDVSGNIVETVQREYVETIDAV  
 RVATFEKLKLNLPDLVNIIEIRYRNSQSLNIALGLALDNLGTGPNSEIVIGAP  
 SMGKTAALMAASNAIKORIPAFESLNSNLILMORLIAEGSVGATNLKGLDL  
 SDFGRIQNAAGEWDAPLYIVDVNMLKILDAVARRLCVGEKIQITFVYIGIIVAD  
 NPAPRYQPAAISQSLKSLARELDIPVALSGVRPAEGSAAPMLADIRSGALEQDA  
 DVNMFHARDNETQTLAKQKRNPIGTVELERQASFTRVCKSP"  
 /complement(1368..1595)  
 /gene="TP0059"  
 /complement(1368..1595)  
 /gene="TP0059"  
 /note="hypothetical protein; identified by Glimmer;  
 putative"  
 /codon\_start=1  
 /transl\_table=1  
 /product="T. pallidum predicted coding region TP0059"  
 /protein\_id="AAC65063.1"  
 /db\_xref="GI:3322326"  
 /db\_xref="GI:3322326"  
 /translation="MRKYLARSMSGCSFSCAKNCATLGLDSTEGRPFLAVGPFPAS  
 LEGRHSGDCLFTTFPPLSTPRRRFCGHA"  
 /complement(1601..2071)  
 /gene="TP0060"  
 /complement(1601..2071)  
 /gene="TP0060"  
 /note="similar to GB:D26185.SP:P37437.PID:467335  
 GB:AL009126 percent identity: 32.67; identified by  
 sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=1  
 /product="ribosomal protein L9 (rpL9)"  
 /protein\_id="AAC65055.1"  
 /db\_xref="GI:3322318"  
 /db\_xref="GI:3322318"  
 /translation="MKITINDVKTIGSGGVKVAAGPKNLYLPRNLVAPNRPY  
 ARFKRQDDIEKRSIKRQDANLKAERADQPYVIANAGNGLCAVTSYTABOL  
 ACMGFEVERKREVPGLTKGVGNHVRILEICAVPVATINQSEADSVSE"  
 /complement(2177..2476)  
 /gene="TP0061"  
 /complement(2177..2476)  
 /gene="TP0061"  
 /note="similar to SP:PI0806 percent identity: 51.95;  
 identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=1

gene  
 CDS

Query Match 80.9%; Score 17.8; DB 3; Length 300994;  
 Best Local Similarity 90.5%; Pred. No. 2.6e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTATCTACATCGAAGCA 22  
 |||||  
 Db 225380 GTATCTACATCGAATCGA 225360

RESULT 9  
 AT093111 2438 bp mRNA linear PLN 21-APR-2002

LOCUS Arabidopsis thaliana putative protein (At3g56140) mRNA, complete  
 DEFINITION cds.

ACCESSION AY093111  
 VERSION AY093111.1 GI:20260423

KEYWORDS F1L.CDN.  
 SOURCE Arabidopsis thaliana.

ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 2438)

REFERENCE Southwick, A., Katlin-Neumann, G., Nguyen, M., Lam, B., Miranda, M.,  
 Palm, C.J., Bowser, L., Jones, T., Banh, J., Carninci, P., Chen, H.,  
 Cheuk, R., Chung, M.K., Hayashizaki, Y., Ishida, J., Kamiya, A.,  
 Kawai, J., Kim, C., Lin, J., Liu, S.X., Narusaka, M., Pham, P.K.,  
 Sakano, H., Sakurai, T., Satou, M., Seki, M., Shinn, P., Yamada, K.,  
 Shinozaki, K., Ecker, J., Theologis, A. and Davis, R.W.  
 Direct Submission

TITLE Submitted (27-MAR-2002) DNA Sequencing and Technology Center,  
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
 USA

COMMENT e-mail for correspondence: arab@sequence.stanford.edu

RIKEN Genomic Sciences Center (GSC) members carried out the  
 collection and clustering of RAFL CDNAS (RAFL CDNA : RIKEN  
 Arabidopsis Full-Length cDNA) : Seki, M., Narusaka, M., Ishida, J.,  
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,  
 Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGSC (SSP) Consortium members carried out the  
 sequencing and annotation of the RAFL CDNAS. Southwick, A.,  
 Nguyen, M., Katlin-Neumann, G., Lam, B., Miranda, M., Palm, C.J.,  
 Bowser, L., Jones, T., Banh, J., Chen, H., Cheuk, R., Chung, M.K.,  
 Kim, C., Lin, J., Liu, S.X., Pham, P.K., Sakano, H., Shinn, P.,  
 Yamada, K., Ecker, J., Theologis, A. and Davis, R.W.

Southwick, A., (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed  
 equally to this work. Shinozaki, K. (RIKEN GSC) and Davis, R.W.  
 (SSP/Stanford) contributed equally to this work as PIs.

FEATURES  
 Location/Qualifiers

source

gene

CDS

1..2438  
 /organism="Arabidopsis thaliana"  
 /db\_xref="taxon:3702"  
 /chromosome="3"  
 /clone="RAFL09-27-C13 (R17974)"  
 /note="This clone is in pBluescript  
 ecotype: Columbia"  
 1..2438  
 /gene="At3g56140"  
 /note="At3g56140"  
 62..2299  
 /gene="At3g56140"  
 /codon\_start=1  
 /product="putative protein"  
 /protein\_id="AA13110.1"  
 /db\_xref="GI:20260424"  
 /translation="MTKNDNFASQPLFHCSSSSRRRRRGFLHPTNLTSEKN  
 NSLSTALSDLPKRTASRAFLAPPLASASLEFKRSVSLASESSSRTYTS  
 ASSAAPPPATTTBPPPPVKKKEETITRTIDATGAPPMAGDKKRWKLLNAR  
 VYLGAEPPPTKDKKELEIVRNLRKCVSEFOI SVALEAPFLDQDDNQYMK  
 RMDGETLKSYTHVMPAORWQVEEPLSTYCRDNSVRLACGPTKLKLVTOAGSIGLS

BASE COUNT 648 a 546 c 556 g 688 t  
 ORIGIN

Query Match 78.2%; Score 17.2; DB 8; Length 2438;  
 Best Local Similarity 86.4%; Pred. No. 6e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGTATCTACATCGAAGCA 22  
 |||||  
 Db 959 CGTAGATCTACATCGAATCGA 980

RESULT 10  
 AT0927282 2919 bp mRNA linear PLN 06-JAN-2002

LOCUS Arabidopsis thaliana partial mRNA for cellular apoptosis

DEFINITION susceptibility protein homologue (CAS gene).

ACCESSION At297282

VERSION At297282.1 GI:18077709

KEYWORDS cas gene; cellular apoptosis susceptibility protein homologue.  
 thale cress.

SOURCE

ORGANISM

Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS

TITLE

Characterisation of CAS from Arabidopsis thaliana as the nuclear  
 export receptor for importin alpha

REFERENCE

AUTHORS

TITLE

Submitted (05-JAN-2001) Haasen D., Institute of Biology II / Cell  
 Biology, University of Freiburg, Schanzlestr. 1, D-79104 Freiburg,  
 GERMANY

FEATURES

source

gene

CDS

1..2916  
 /organism="Arabidopsis thaliana"  
 /variety="ecotype Col-0"  
 /db\_xref="taxon:3702"  
 1..2916  
 /gene="CAS"  
 <1..2916  
 /gene="CAS"  
 /function="nuclear export receptor for importin alpha"  
 /note="CAS interacts with importin alpha 1-4 and with ran  
 from Arabidopsis thaliana"  
 /codon\_start=1  
 /evidence=experimental  
 /product="cellular apoptosis susceptibility protein  
 homologue"  
 /protein\_id="CAC83300.1"  
 /db\_xref="GI:18077710"  
 /translation="MEMNRRTLVLSQCFNTLSPDPPRTAERALSADAQNYGL  
 AVLRIVAPALDEQTRHAAVNFKNHLSRHPAGDSGISIVSEKEQITLVLSLM  
 LSASPRISQSLSEALTVGKDFPRAMPALPELIANTONALAGDYVSNGIIGTAS  
 SIFKFSYEYRTDALFVLDKCYCLDNFAPLPEIFLKTSLSDSASGSGSPILKPAF  
 ESORLCCTIEYSINPQDLPEFEDHMKEMGEKFKYSSNPALSTEEGILVDDLR  
 AALGENTNHYLEKNEDEBROGLNIFASVMTLLDYSKSPSRLATTEGILVDSVT  
 SVHHLRFGDVAIVEICOSTYIPVNSLVEDEEETEMNTIEFTRDMSGDYDRRTI  
 ACCLTKGLDITQNFQVTEVNSLEDELDSRVNSPMAKASLKLTLRSHIRPFPMQKE  
 GASVSTGIDVQNFQVTEVNSLEDELDSRVNSPMAKASLKLTLRSHIRPFPMQKE  
 PELVRFLEAESNVVHVAASCIIEKLILVKEGARGNRYAGDLSPLFLQMTNFDAL

AUTHORS Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.  
TITLE Direct Submission  
JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA

FEATURES  
source Location/Qualifiers  
1. 300994

gene  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="X"  
complement(<16198..>18483)

mRNA  
/gene="CG8527"  
/map="16B4-16B6"  
/db\_xref="FLYBASE:FBan0008527"  
/db\_xref="FLYBASE:FBgn0030844"  
complement(join(<16198..16317,16386..16574,16635..16845,  
16915..17055,17129..17411,17526..17607,17676..17975,  
18107..>18483))

CDS  
/gene="CG8527"  
/product="CT24901"  
/db\_xref="FLYBASE:FBan0008527"  
/db\_xref="FLYBASE:FBgn0030844"  
complement(join(16198..16317,16386..16574,16635..16845,  
16915..17055,17129..17411,17526..17607,17676..17975,  
18107..18483))

/note="CG8527 gene product"  
/codon\_start=1  
/protein\_id="AAF48734.1"  
/db\_xref="GI:7293356"  
/db\_xref="FLYBASE:FBan0008527"  
/db\_xref="FLYBASE:FBgn0030844"  
/translation="MPQEKRHPESTQFLETLVIFRSLSLYQTKKEFQNSTLHGVR  
YLAESGRPIGEMWECFTSGAVTALVIMSMEKQNPITGLDIFHQNVYF  
TVYQCEAADFDEKTYEKYNTLANDEAQMYPFLILSLNEFNDPAKVSOS  
IPONLDARTIREAFEGHIDCKNVFSCREDIPCCDHEPITYETENGFPYAFNR  
FKSTPEDVKTGAHDLYETDKKWFLEFIPNSTRFETISNEEYSGSDPNADISBP  
QLYEATISKNTTITTDAROLSGQKCFSDPYKNTFPDATTSSCKQKQKANKAI  
KTCNCPFTKPLPNVPCSIKDFDLDFKASITINIKDLOCLELSCSTVNIIDLI  
KMSDRESLGVLEFLTWPIIRKREVLGWDLVSGIASLAFGLFLGVEIY  
YFLRACQVYKNRQELVEEIKRIQEPKPIDLKLSKSNPRIDGPTPSAVLYKPP  
AESAQPNPLGQYQRHEKDYSLNTKNYKTKPVVDGYTSTKTKNSMTYDHSQMP  
"

gene  
complement(<21273..>28555)  
/gene="CG8465"  
/map="16B4-16B7"  
/db\_xref="FLYBASE:FBan0008465"  
/db\_xref="FLYBASE:FBgn0030845"  
complement(join(21273..21792,21868..21999,22058..22191,  
22257..22823,22889..23887,24168..24320,24398..25116,  
25191..25370,27861..28115,28453..28555))

mRNA  
/gene="CG8465"  
/product="CT24733"  
/db\_xref="FLYBASE:FBan0008465"  
/db\_xref="FLYBASE:FBgn0030845"  
complement(join(21273..21792,21868..21999,22058..22191,  
22257..22823,22889..23887,24168..24320,24398..25116,  
25191..25370,27861..28246))

mRNA  
/gene="CG8465"  
/product="CT24757"  
/db\_xref="FLYBASE:FBan0008465"  
/db\_xref="FLYBASE:FBgn0030845"  
complement(join(21273..21792,21868..21999,22058..22191,  
22257..22823,22889..23887,24168..24320,24398..25116,  
25191..25448))

CDS  
/gene="CG8465"  
/product="CT24743"  
/db\_xref="FLYBASE:FBan0008465"  
/db\_xref="FLYBASE:FBgn0030845"  
complement(join(21399..21792,21868..21999,22058..22191,  
22257..22823,22889..23887,24168..24320,24398..25116,  
25191..25370,27861..28092))  
/note="CG8465 gene product [alt 1]"

/codon\_start=1  
/protein\_id="AAF48735.1"  
/db\_xref="GI:7293357"  
/db\_xref="FLYBASE:FBan0008465"  
/db\_xref="FLYBASE:FBgn0030845"  
/translation="MSTFYGYIPSKAGCEGVSQCSIGIAAVNIKPSNPASGAS  
VASGSPSGSASVQGNADDSAAATKYDDPPDPSPMLIFTEKSKALDILRYKKA  
RLREPNLEASQESVYQFESIEMALKRCKAPSPKPIPIISGGYKSPSTNSCS  
SSPTNGSGSFPILPIGSSNSMSNILLSDSPSSPSNSVIVANGQOQOQOQOQO  
PDVSEGPFPAPATKQELVEFRKQIEGHIDRVKRIITENRPIISSGDTPTISKEC  
RIMAHICAOYNKARIQILLKTIISDFEITOLYVKGSGKMCALNISLIDYLINP  
DKRGETPLHFAKNGHVAWEVLYSEYCKSLNHEKEKEKEIICRNANAHTVTK  
KIELLYDPHFVPLYSQNTLPRVQGFSPSKBPNTQHNADYDEGLSVLAISALA  
GPMSEKAMNFRWKTPPPRVNNMSESLASPPSKIVPVSISIFDSRGNSEPVH  
SGRRLPLLEATSSPKPTKNVNMGTTECHNNNNKVPVPLEFPATPIRKMPDLF  
MAYRNNSEFDSPLADSOIIDMSISRLNSLSDPFEERKIKNTDIEKGLVYGRDL  
AROEDLEKREYWDPLDSFLDIGTEGLRLAFLVLEKTEBOQADSSEYWNNAHJOYE  
DSMAEQOQOIRKQKNEAGATSPSAGYMPYTOVEKSIQVAFARITKTLINKNVY  
SINDLLELRLSLIVSRDARFISVDPSKSHRAHVAHVHISQSVYAML  
OLIOMLRLROLADERGEOHLGCVCASLLMLEOAPTSAVHLPTLKEBELCAAM  
ETEOCCALMDANLSRKTSSRRKRLSLAAVVOQLODTSGTSSSALHSLGVA  
STSIGASRVASASKDMARQOODEDEYQIFPCTVTLTPYGSSEDEENFT  
POSTSGTISMDLEPRYELTFQNEPPTKRDIDVNAISNDIDKPTLPHYAKTAME  
SYSCAEMLNKVKQPEPWYGTSSSHNSQTLPPKRLATPKLNAVSGRSGPIL  
APVPELRATPSASIOVASSETNGSEVTAVPASPIISFALTAQAQSFOTPLNKVR  
GLFSYRQRSYNEGDPLEGNRN"  
complement(join(21399..21792,21868..21999,22058..22191,  
22257..22823,22889..23887,24168..24320,24398..25116,  
25191..25370,27861..28092))  
/gene="CG8465"  
/note="CG8465 gene product [alt 2]"  
/codon\_start=1  
/protein\_id="AAF48736.1"  
/db\_xref="GI:7293358"  
/db\_xref="FLYBASE:FBan0008465"  
/db\_xref="FLYBASE:FBgn0030845"  
/translation="MSTFYGYIPSKAGCEGVSQCSIGIAAVNIKPSNPASGAS  
VASGSPSGSASVQGNADDSAAATKYDDPPDPSPMLIFTEKSKALDILRYKKA  
RLREPNLEASQESVYQFESIEMALKRCKAPSPKPIPIISGGYKSPSTNSCS  
SSPTNGSGSFPILPIGSSNSMSNILLSDSPSSPSNSVIVANGQOQOQOQOQO  
PDVSEGPFPAPATKQELVEFRKQIEGHIDRVKRIITENRPIISSGDTPTISKEC  
RIMAHICAOYNKARIQILLKTIISDFEITOLYVKGSGKMCALNISLIDYLINP  
DKRGETPLHFAKNGHVAWEVLYSEYCKSLNHEKEKEKEIICRNANAHTVTK  
KIELLYDPHFVPLYSQNTLPRVQGFSPSKBPNTQHNADYDEGLSVLAISALA  
GPMSEKAMNFRWKTPPPRVNNMSESLASPPSKIVPVSISIFDSRGNSEPVH  
SGRRLPLLEATSSPKPTKNVNMGTTECHNNNNKVPVPLEFPATPIRKMPDLF  
MAYRNNSEFDSPLADSOIIDMSISRLNSLSDPFEERKIKNTDIEKGLVYGRDL  
AROEDLEKREYWDPLDSFLDIGTEGLRLAFLVLEKTEBOQADSSEYWNNAHJOYE  
DSMAEQOQOIRKQKNEAGATSPSAGYMPYTOVEKSIQVAFARITKTLINKNVY  
SINDLLELRLSLIVSRDARFISVDPSKSHRAHVAHVHISQSVYAML  
OLIOMLRLROLADERGEOHLGCVCASLLMLEOAPTSAVHLPTLKEBELCAAM  
ETEOCCALMDANLSRKTSSRRKRLSLAAVVOQLODTSGTSSSALHSLGVA  
STSIGASRVASASKDMARQOODEDEYQIFPCTVTLTPYGSSEDEENFT  
POSTSGTISMDLEPRYELTFQNEPPTKRDIDVNAISNDIDKPTLPHYAKTAME  
SYSCAEMLNKVKQPEPWYGTSSSHNSQTLPPKRLATPKLNAVSGRSGPIL  
APVPELRATPSASIOVASSETNGSEVTAVPASPIISFALTAQAQSFOTPLNKVR  
GLFSYRQRSYNEGDPLEGNRN"

CDS  
complement(join(21399..21792,21868..21999,22058..22191,  
22257..22823,22889..23887,24168..24320,24398..25116,  
25191..25410))  
/gene="CG8465"  
/note="CG8465 gene product [alt 3]"  
/codon\_start=1  
/protein\_id="AAF48737.1"  
/db\_xref="GI:7293359"  
/db\_xref="FLYBASE:FBan0008465"  
/db\_xref="FLYBASE:FBgn0030845"  
/translation="MPTQHCYHRHGDADSPMLIFTEKSKALDILRYKKA  
NLEQESVVOGFESEIEMALKRCKAPSPKPIPIISGGYKSPSTNSCS  
SSPTNGSGSFPILPIGSSNSMSNILLSDSPSSPSNSVIVANGQOQOQOQOQO  
PDVSEGPFPAPATKQELVEFRKQIEGHIDRVKRIITENRPIISSGDTPTISKEC  
RIMAHICAOYNKARIQILLKTIISDFEITOLYVKGSGKMCALNISLIDYLINPDKRG



Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schaefer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (08-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
(bases 1 to 144662)

3 (bases 1 to 144662)  
Britten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campobasso, A., Chang, J., Chazaro, B., Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Coyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schaefer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

## COMMENT

Submitted (02-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 2, 2002 this sequence version replaced g1:18308604.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Web site: <http://www.seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: 2005\_H.15

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 134421: contig of 134421 bp in length  
\* 134422 134521: gap of 100 bp  
\* 134522 144662: contig of 10141 bp in length.

## FEATURES

## Source

1. 144662  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15"  
/clone="CTD-2005H15"

## BASE COUNT

42042 a 30377 c 29991 g 42108 t 144 others

## ORIGIN

Query Match 80.9% Score 17.8; DB 2; Length 144662;

Best Local Similarity 90.5% Pred No. 2.7e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 118383 GTCATCTACATTCGAAACGA 118403

2 GTCATCTACATTCGAAACGA 22

11 ||||| |||||

AC012162 169856 bp DNA linear INV 05-MAY-2001

AC012162 Drosophila melanogaster, chromosome X, region 16D-16F, BAC clone

AC012162 BACROIN10, complete sequence.

AC012162.10 GI:13957596

HTG.

Drosophila melanogaster.

Drosophila melanogaster.

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 169856)

Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H.,

Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,

Rogers, Y., An, H., Baldwin, D., Bonzon, J., Beeson, K.Y., Busam, D.A.,

Carlson, J.W., Center, A., Chapple, M., Davenport, L.B., Dietz, S.M.,

Dotson, J.W., Dorset, V., Dou, L.E., Doyle, C., Dresnek, D., Fartan, D.,

Ferrera, S., Flise, E., Galle, R.F., Garg, N.S., George, R.A.,

Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, J.J.,

Idegian, C., Jaitai, M., Kruse, D., Li, P., Matelli, B., Moshirefi, A.,

McInosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,

Pachlob, J., Pargay, V., Park, S., Patel, S., Pfeiffer, B.,

Phonemavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,

Stapleton, M., Strong, R., Svirska, R., Tector, C., Williams, S.M.,

Zaveri, V.S., Smith, H.O., Rubin, G.M., and Venter, J.C.

Sequencing of Drosophila chromosome X, region 16D-16F

Unpublished

2 (bases 1 to 169856)

Celniker, S.E., Agayanni, A., Arcina, T.T., Baxter, E., Blazef, R.G.,

Butenhoff, C., Chapple, M., Chavez, C., Chew, M., Ciesiolka, L.,

Doyle, C.M., Fartan, D.E., Galle, R., George, R.A., Harris, N.L.,

Hoskins, R.A., Houston, K.A., Hummel, S.R., Katta, K., Kearney, L.,

Kim, E., Lee, B., Lewis, S., Li, P., Lomocan, M.A., Mazza, P.,

Moshirefi, A.R., Moshirefi, M., Nixon, K., Pachlob, J.M., Park, S.,

Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,

Svirska, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and

Rubin, G.M.

Submitted (21-OCT-1999) Drosophila Genome Center, Lawrence Berkeley

Laboratory, MS 64-121, Berkeley, CA 94720, USA

On May 5, 2001 this sequence version replaced g1:6957580.

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome

shotgun and from subclones of this BAC and its neighboring clones.

For further information about this sequence, including its location

and relationship to other sequences, please visit our sequence

archive Web site (<http://www.fruitfly.org/sequence/>) or send email

to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).

Location/Qualifiers

1. 169856

/organism="Drosophila melanogaster"

/strain="Y; cn bw sp"

/db\_xref="taxon:7227"

/chromosome="X"

/map="16D-16F"

/clone="BACROIN10 (D1115)"

/clone\_lib="RPCI-98 (Rosalind Park Cancer Institute

Drosophila melanogaster BAC library, partial EcoRI in

```

repeat_region 12549 . 15363

Query Match      80.9%; Score 17.8; DB 9; Length 127682;
Best Local Similarity 90.5%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTATACCTACATTCGAAACGA 22
|||||
Db 60288 GATATCTACATTCACAAACGA 60268

RESULT 4
AC129980/ 132202 bp DNA linear HNG 19-AUG-2002
LOCUS Homo sapiens chromosome 15 clone CTD-2529M12 map 15, *** SEQUENCING
DEFINITION IN PROGRESS ***, 2 ordered pieces.
ACCESSION AC129980.3 GI:229297424
VERSION HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVERIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 132202)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
JOURNAL Unpublished
TITLE Homo sapiens chromosome 15, clone CTD-2529M12
2 (bases 1 to 132202)
REFERENCE 2 (bases 1 to 132202)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhallier,B.,
Camata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
PhunKhang,P., Pierre,N., Raymond,S., Schuppach,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
JOURNAL Direct Submission
TITLE Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT 3 (bases 1 to 132202)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhallier,B.,
Camata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
PhunKhang,P., Pierre,N., Raymond,S., Schuppach,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
JOURNAL Direct Submission
TITLE Submitted (19-AUG-2002) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Aug 19, 2002 this sequence version replaced g1:2213335.
All repeats were identified using RepeatMasker:

```

```

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WILR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L22204
Center clone name: 2529_M12
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 128: contig of 128 bp in length
* 129 228: gap of 100 bp
* 229 132202: contig of 131974 bp in length.
* Location/Qualifiers
  1..132202
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="15"
    /map="15"
    /clone="CTD-2529M12"
    /clone_lib="CTD2 Human BAC"
BASE COUNT 37938 a 27502 c 28043 g 38577 t 142 others
ORIGIN

Query Match      80.9%; Score 17.8; DB 2; Length 132202;
Best Local Similarity 90.5%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTATACCTACATTCGAAACGA 22
|||||
Db 16235 GTCTACCTACATTCGAAACGA 16215

RESULT 5
AC104374 144662 bp DNA linear HNG 02-JUN-2002
LOCUS Homo sapiens chromosome 15 clone CTD-2005H15 map 15, *** SEQUENCING
DEFINITION IN PROGRESS ***, 2 ordered pieces.
ACCESSION AC104374.3 GI:21313764
VERSION HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVERIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 144662)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhallier,B.,
Brown,A., Camata,J., Campioano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Gardyna,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Labouque,K.,
Lamasares,R., Landers,T., Lehocsky,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPeckers,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,

```

```
AluY repeat: matches 75. .310 of consensus"
repeat_region 239. .538
/note="LIM4 repeat: matches 2107. .2438 of consensus"
repeat_region 336. .538
/note="LIM4 repeat: matches 2227. .2438 of consensus"
repeat_region 550. .153
/note="LIM4 repeat: matches 2582. .3669 of consensus"
misc_feature complement(565. .1283)
/note="match: GSS: Em:AQ383567"
675. .1155
/note="LIM4 repeat: matches 2722. .3226 of consensus"
misc_feature complement(698. .1442)
/note="match: GSS: Em:AQ896860"
misc_feature complement(906. .1442)
/note="match: GSS: Em:AQ761517"
1288. .1539
/note="match: GSS: Em:AQ879547"
1290. .1693
/note="match: GSS: Em:AQ826107"
1294. .2055
/note="match: GSS: Em:AQ901298"
1294. .1860
/note="match: GSS: Em:AQ047100"
1362. .2055
/note="match: GSS: Em:AQ490007"
1362. .1961
/note="match: GSS: Em:AQ585548"
1362. .1761
/note="match: GSS: Em:AQ275646"
1447. .1837
/note="match: GSS: Em:AQ355624"
1448. .1964
/note="match: GSS: Em:AQ474420"
1449. .2054
/note="match: GSS: Em:AQ76221"
1450. .2017
/note="match: GSS: Em:AQ537460"
1451. .2223
/note="match: GSS: Em:AQ738677"
1451. .2150
/note="match: GSS: Em:AQ738952"
1451. .1896
/note="match: GSS: Em:AQ669827"
1451. .1877
/note="match: GSS: Em:AQ810603"
1451. .1827
/note="match: GSS: Em:AQ718640"
1466. .1877
/note="match: GSS: Em:AQ692785"
1592. .2208
/note="LIMB1 repeat: matches 5338. .5979 of consensus"
LIMB1 repeat: matches 5338. .5979 of consensus"
2199. .2404
/note="LIM4 repeat: matches 3935. .4135 of consensus"
LIM4 repeat: matches 3935. .4135 of consensus"
complement(2626. .3066)
/note="match: GSS: Em:AQ245895"
complement(2887. .3061)
/note="match: GSS: Em:B91023"
3067. .3302
/note="match: GSS: Em:AQ009565"
3074. .3453
/note="match: GSS: Em:AQ261089"
3074. .3302
/note="match: GSS: Em:AQ059372"
3407. .3448
/note="MERB repeat: matches 1. .42 of consensus"
complement(3620. .4114)
/note="match: GSS: Em:AQ117010"
complement(3956. .4111)
/note="match: GSS: Em:AQ304953"
4136. .4803
/note="match: GSS: Em:AQ531595"

repeat_region 4636. .4942
/note="AluSc repeat: matches 1. .308 of consensus"
AluSx repeat: matches 1. .311 of consensus"
misc_feature 5143. .3591
/note="match: GSS: Em:AQ224449"
5147. .5738
/note="match: GSS: Em:B60185"
5147. .5659
/note="match: GSS: Em:AQ076436"
5154. .3605
/note="match: GSS: Em:AQ608022"
5204. .5675
/note="match: GSS: Em:AQ712434"
5702. .5976
/note="LTR16A repeat: matches 132. .428 of consensus"
LTR16A repeat: matches 132. .428 of consensus"
6212. .6878
/note="match: GSS: Em:AQ321209"
6212. .6775
/note="match: GSS: Em:AQ377996"
6213. .6543
/note="match: GSS: Em:AQ382827"
6216. .6770
/note="match: GSS: Em:AQ344327"
6221. .6541
/note="match: GSS: Em:AQ052109"
6229. .6703
/note="match: GSS: Em:AQ370102"
6254. .6541
/note="match: GSS: Em:AQ214603"
6256. .6733
/note="match: GSS: Em:AQ454556"
6272. .6541
/note="match: GSS: Em:AQ095672"
6329. .6686
/note="match: GSS: Em:AQ629712"
6397. .6873
/note="match: GSS: Em:AQ587044"
6406. .6787
/note="match: GSS: Em:AQ587048"
6546. .6888
/note="match: GSS: Em:AQ052109"
6546. .6741
/note="match: GSS: Em:AQ095672"
6546. .6739
/note="match: GSS: Em:AQ214603"
7281. .7429
/note="MER65A repeat: matches 295. .445 of consensus"
MER65A repeat: matches 295. .445 of consensus"
7454. .7774
/note="MER65A repeat: matches 15. .334 of consensus"
MER65A repeat: matches 15. .334 of consensus"
8040. .8082
/note="MABE1 repeat: matches 34. .77 of consensus"
MABE1 repeat: matches 34. .77 of consensus"
8437. .8643
/note="LIPB1 repeat: matches 6008. .6155 of consensus"
LIPB1 repeat: matches 6008. .6155 of consensus"
8933. .9240
/note="LTR40a repeat: matches 2. .327 of consensus"
LTR40a repeat: matches 2. .327 of consensus"
9246. .10036
/note="LI repeat: matches 4060. .4853 of consensus"
LI repeat: matches 4060. .4853 of consensus"
10042. .10325
/note="LTR14a repeat: matches 1. .344 of consensus"
LTR14a repeat: matches 1. .344 of consensus"
10326. .12101
/note="HERVK4 repeat: matches 1. .1888 of consensus"
HERVK4 repeat: matches 1. .1888 of consensus"
12102. .12546
/note="MER9 repeat: matches 1. .509 of consensus"
MER9 repeat: matches 1. .509 of consensus"
```



```

* 22588 23270: contig of 683 bp in length
* 23271 23370: gap of 100 bp
* 23371 24068: contig of 698 bp in length
* 24069 24168: gap of 100 bp
* 24169 24855: contig of 687 bp in length
* 24856 24955: gap of 100 bp
* 24956 25652: contig of 697 bp in length
* 25653 25752: gap of 100 bp
* 25753 26429: contig of 677 bp in length
* 26430 26529: gap of 100 bp
* 26530 27231: contig of 702 bp in length
* 27232 27331: gap of 100 bp
* 27332 28045: contig of 714 bp in length
* 28046 28145: gap of 100 bp
* 28146 28868: contig of 723 bp in length
* 28869 28968: gap of 100 bp
* 28969 29688: contig of 720 bp in length
* 29689 29788: gap of 100 bp
* 29789 30507: contig of 719 bp in length
* 30508 30607: gap of 100 bp
* 30608 31313: contig of 706 bp in length
* 31314 31413: gap of 100 bp
* 31414 32105: contig of 692 bp in length
* 32106 32205: gap of 100 bp
* 32206 32900: contig of 695 bp in length
* 32901 33000: gap of 100 bp
* 33001 33684: contig of 684 bp in length
* 33685 33784: gap of 100 bp
* 33785 34486: contig of 702 bp in length
* 34487 34586: gap of 100 bp
* 34587 35310: contig of 724 bp in length
* 35311 35410: gap of 100 bp
* 35411 36129: contig of 719 bp in length
* 36130 36229: gap of 100 bp
* 36230 36944: contig of 715 bp in length
* 36945 37044: gap of 100 bp
* 37045 37759: contig of 715 bp in length
* 37760 37859: gap of 100 bp
* 37860 38552: contig of 693 bp in length
* 38553 38652: gap of 100 bp
* 38653 39374: contig of 722 bp in length
* 39375 39474: gap of 100 bp
* 39475 40190: contig of 716 bp in length
* 40191 40290: gap of 100 bp
* 40291 41008: contig of 718 bp in length
* 41009 41108: gap of 100 bp
* 41109 41800: contig of 692 bp in length
* 41801 41900: gap of 100 bp
* 41901 42597: contig of 697 bp in length
* 42598 42697: gap of 100 bp
* 42698 43390: contig of 693 bp in length
* 43391 43490: gap of 100 bp
* 43491 44186: contig of 696 bp in length
* 44187 44286: gap of 100 bp
* 44287 44496: contig of 710 bp in length
* 44957 45056: gap of 100 bp
* 45057 45815: contig of 719 bp in length
* 45816 45915: gap of 100 bp
* 45916 46631: contig of 716 bp in length
* 46632 46731: gap of 100 bp
* 46732 47448: contig of 718 bp in length
* 47450 47549: gap of 100 bp
* 47550 48275: contig of 726 bp in length
* 48276 48375: gap of 100 bp
* 48376 49086: contig of 711 bp in length
* 49087 49186: gap of 100 bp
* 49187 49910: contig of 724 bp in length
* 49911 50010: gap of 100 bp
* 50011 50723: contig of 713 bp in length
* 50724 50823: gap of 100 bp
* 50824 51512: contig of 689 bp in length
* 51513 51612: gap of 100 bp
* 51613 52318: contig of 706 bp in length

```

```

Query Match      60.9%; Score 17.8; DB 2; Length 72579;
Best Local Similarity 90.5%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY      2  GTATCTCATTCGAAACGA 22
Db      43249  GTCTATCTCATTCGAAACGA 43269

```

```

RESULT 3
AL137861/c
LOCUS
DEFINITION
Human DNA sequence from clone RP4-813B7 on chromosome 1. Contains a
pseudogene similar to zinc finger protein ZNF186, STS and GSSS.

```

```

ACCESSION      AL137861
VERSION        AL137861
KEYWORDS       complete sequence.
SOURCE         human.
ORGANISM       Homo sapiens

```

```

REFERENCE      Eukariyola; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE          1 (bases 1 to 127682)
JOURNAL        Direct Submission

```

```

COMMENT        Submitted (18-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
                CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

```

On Jul 14, 2000 this sequence version replaced gi:18894204. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> RP4-813B7 is from the library RPCI-4 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTOR: pCYPAC2

This sequence is the entire insert of clone RP4-813B7. Location/Qualifiers

```

FEATURES
source

```

```

1..127682
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP4-813B7"
/clone_lib="RPCI-4"
1..235
/note="AluY repeat: matches 75..310 of consensus
repeat_region

```

TITLE Direct Submission  
JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA  
COMMENT This sequence was identified as CDM:10210993 by the submitter. For more information on this record e-mail to fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
FEATURES  
1. 20306  
Location/Qualifiers  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
BASE COUNT 5977 a 4332 c 4420 g 5577 t  
ORIGIN  
Query Match 80.9%; Score 17.8; DB 2; Length 20306;  
Best Local Similarity 90.5%; Pred. No. 2.8e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 GATATCTACATTCGAACGA 22  
16836 GATATCTACATTCGAACGA 16836  
Db  
RESULT 2  
AC103875 72579 bp DNA linear HTG 29-NOV-2001  
LOCUS AC103875  
DEFINITION Homo sapiens chromosome 11 clone CTD-2010D2 map 11, LOW-PASS  
AC103875  
SEQUENCE SAMPLING.  
AC103875.1 GI:17149760  
HTG; HTGS\_PHASE0.  
SOURCE  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 72579)  
Biren, B., Linton, L., Nusbaum, C., and Lander, E.  
Homo sapiens chromosome 11, clone CTD-2010D2  
2 (bases 1 to 72579)  
Unpublished  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
1 (bases 1 to 72579)  
Biren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barta, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B., Choquel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, D., Dodge, S., Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamarez, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., MacDonald, P., Major, J., Marguis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPeckers, R., Meldrum, J., Menes, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Piere, N., Pollara, Y., Raymond, C., Retter, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, N., Schunback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testa, S., Theodores, J., Topham, K., Travers, M., Travis, N., Trifillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
TITLE  
JOURNAL  
COMMENT  
Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
Center: Whitehead Institute/MIT Center for Genome Research  
Center code: WIR

Web site: <http://www.seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L22053  
Center clone name: 2010\_D\_2  
-----  
\* NOTE: This record contains 90 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
1 705: contig of 705 bp in length  
\* 706 805: gap of 100 bp  
\* 806 1528: contig of 723 bp in length  
\* 1529 1628: gap of 100 bp  
\* 1629 2350: contig of 722 bp in length  
\* 2351 2450: gap of 100 bp  
\* 2451 3163: contig of 713 bp in length  
\* 3164 3263: gap of 100 bp  
\* 3264 3990: contig of 727 bp in length  
\* 3991 4090: gap of 100 bp  
\* 4091 4797: contig of 707 bp in length  
\* 4798 4897: gap of 100 bp  
\* 4898 5589: contig of 692 bp in length  
\* 5590 5689: gap of 100 bp  
\* 5690 6395: contig of 706 bp in length  
\* 6396 6495: gap of 100 bp  
\* 6496 7182: contig of 687 bp in length  
\* 7183 7282: gap of 100 bp  
\* 7283 7965: contig of 683 bp in length  
\* 7966 8065: gap of 100 bp  
\* 8066 8781: contig of 716 bp in length  
\* 8782 8881: gap of 100 bp  
\* 8882 9562: contig of 681 bp in length  
\* 9563 9662: gap of 100 bp  
\* 9663 10388: contig of 726 bp in length  
\* 10389 10488: gap of 100 bp  
\* 10489 11210: contig of 722 bp in length  
\* 11211 11310: gap of 100 bp  
\* 11311 12031: contig of 721 bp in length  
\* 12032 12331: gap of 100 bp  
\* 12332 12834: contig of 703 bp in length  
\* 12835 12934: gap of 100 bp  
\* 12935 13626: contig of 692 bp in length  
\* 13627 13726: gap of 100 bp  
\* 13727 14427: contig of 701 bp in length  
\* 14428 14527: gap of 100 bp  
\* 14528 15228: contig of 701 bp in length  
\* 15229 15328: gap of 100 bp  
\* 15329 16014: contig of 686 bp in length  
\* 16015 16114: gap of 100 bp  
\* 16115 16828: contig of 714 bp in length  
\* 16829 16928: gap of 100 bp  
\* 16929 17650: contig of 722 bp in length  
\* 17651 17750: gap of 100 bp  
\* 17751 18454: contig of 704 bp in length  
\* 18455 18554: gap of 100 bp  
\* 18555 19218: contig of 664 bp in length  
\* 19219 19318: gap of 100 bp  
\* 19319 20026: contig of 708 bp in length  
\* 20027 20126: gap of 100 bp  
\* 20127 20850: contig of 724 bp in length  
\* 20851 20950: gap of 100 bp  
\* 20951 21668: contig of 718 bp in length  
\* 21669 21768: gap of 100 bp  
\* 21769 22487: contig of 719 bp in length  
\* 22488 22587: gap of 100 bp

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:07:09 ; Search time 78.1344 Seconds  
(without alignments)  
8194.363 Million cell updates/sec

Title: US-09-477-082-30

Perfect score: 22

Sequence: 1 cgtatctacatcgaacga 22

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_ph:\*  
7: gb\_pl:\*  
8: gb\_pr:\*  
9: gb\_ro:\*  
10: gb\_sts:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_vl:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_un:\*  
28: em\_vl:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pin:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.8	80.9	20306	2 AC019801	AC019801 Drosophila
2	17.8	80.9	72379	2 AC019875	AC019875 Homo sapi
3	17.8	80.9	127682	2 AL137861	AL137861 Human DNA
4	17.8	80.9	132202	2 AC0129980	AC0129980 Homo sapi
5	17.8	80.9	144662	2 AC0104374	AC0104374 Homo sapi
6	17.8	80.9	169856	3 AC012162	AC012162 Drosophila
7	17.8	80.9	173281	3 AC012161	AC012161 Drosophila
8	17.8	80.9	300994	3 AE003506	AE003506 Drosophila
9	17.2	78.2	2438	8 AY093111	AY093111 Arabidops
10	17.2	78.2	2919	8 ATH297282	AJ297282 Arabidops
11	17.2	78.2	9289	6 AX251255	AX251255 Sequence
12	17.2	78.2	11101	1 AE001191	AE001191 Treponema
13	17.2	78.2	82454	8 AC012193	AC012193 Arabidops
14	17.2	78.2	84825	8 AC006418	AC006418 Arabidops
15	17.2	78.2	95214	8 AC006526	AC006526 Arabidops
16	17.2	78.2	99492	8 ATF18021	AL163763 Arabidops
17	17.2	78.2	128739	2 AC016602	AC016602 Mus muscu
18	17.2	78.2	197662	2 AC099713	AC099713 Mus muscu
19	17.2	78.2	215126	2 AC068316	AC068316 Mus muscu
20	17.2	78.2	223580	2 AC079216	AC079216 Mus muscu
21	17.2	78.2	248201	2 AC015293	AC015293 Mus muscu
22	16.8	76.4	1501	10 AF244366	AF244366 Rattus no
23	16.8	76.4	3444	1 BAN249463	AE013765 Bacillus
24	16.8	76.4	11108	1 AE013765	AE013765 Versinia
25	16.8	76.4	27559	8 SC8339	249210 S. cerevisia
26	16.8	76.4	58427	8 ATF1515	AL101023 Arabidops
27	16.8	76.4	100310	2 F17A13	AL096692 Arabidops
28	16.8	76.4	133256	2 AC127021	AC127021 Medicago
29	16.8	76.4	153382	9 HS111B22	298200 Human DNA s
30	16.8	76.4	157375	2 AC122095	AC122095 Rattus no
31	16.8	76.4	157903	9 AP002961	AP002961 Homo sapi
32	16.8	76.4	159635	2 AC011009	AC011009 Homo sapi
33	16.8	76.4	159705	2 AC021007	AC021007 Homo sapi
34	16.8	76.4	161687	9 AC046351	AC046351 Homo sapi
35	16.8	76.4	174499	2 AC066694	AC066694 Homo sapi
36	16.8	76.4	175179	2 AC072025	AC072025 Homo sapi
37	16.8	76.4	175977	2 AC058808	AC058808 Homo sapi
38	16.8	76.4	192155	2 AC123462	AC123462 Rattus no
39	16.8	76.4	194143	8 ATCHRIV48	AL161548 Arabidops
40	16.8	76.4	200001	2 ATCHRIV71	AL161575 Arabidops
41	16.8	76.4	204993	2 AC024447	AC024447 Homo sapi
42	16.8	76.4	205158	2 AC068980	AC068980 Homo sapi
43	16.8	76.4	256674	2 AC023594	AC023594 Homo sapi
44	16.8	76.4	288652	2 AC092942	AC092942 Homo sapi
45	16.8	76.4	313050	1 AJ414152	AJ414152 Versinia

## ALIGNMENTS

RESULT 1  
AC019801  
LOCUS AC019801  
DEFINITION AC019801 20306 bp DNA linear HTG 03-JAN-2000  
ACCESSION AC019801  
VERSION AC019801.1 GI:6665096  
KEYWORDS HTG; HTGS; PHASE2.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 20306)  
AUTHORS Adams, M. and Venter, J. C.

NUMBER OF SEQ ID NOS: 8603  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO: 563  
 LENGTH: 4048  
 TYPE: DNA  
 ORGANISM: Aspergillus fumigatus  
 US-10-128-714-563

Query Match 72.4%; Score 15.2; DB 9; Length 4048;  
 Best Local Similarity 85.0%; Pred. No. 2.7e+02;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGATTGCGAGATTGCG 20  
 DB 1016 TCGGGGATTGACACATTGCG 997

## RESULT 13

US-10-128-714-5563/c  
 Sequence 5563, Application US/10128714  
 Publication No. US20030119013A1  
 GENERAL INFORMATION:  
 APPLICANT: Jiang, Bo  
 APPLICANT: Hu, Wengqi  
 APPLICANT: Tishkoff, Daniel  
 APPLICANT: Zamudio, Carlos  
 APPLICANT: Eroshtkin, Alexey M  
 APPLICANT: Lemieux, Sebastien M  
 TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
 FILE REFERENCE: 10182-018-999  
 CURRENT APPLICATION NUMBER: US/10-128-714  
 CURRENT FILING DATE: 2002-04-23  
 PRIOR APPLICATION NUMBER: US 60/285,697  
 PRIOR FILING DATE: 2001-04-23  
 PRIOR APPLICATION NUMBER: US 60/287,066  
 PRIOR FILING DATE: 2001-04-27  
 PRIOR APPLICATION NUMBER: US 60/295,890  
 PRIOR FILING DATE: 2001-06-05  
 PRIOR APPLICATION NUMBER: US 60/303,899  
 PRIOR FILING DATE: 2001-07-09  
 PRIOR APPLICATION NUMBER: US 60/316,362  
 PRIOR FILING DATE: 2001-08-31  
 NUMBER OF SEQ ID NOS: 8603  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO: 5563  
 LENGTH: 4048  
 TYPE: DNA  
 ORGANISM: Aspergillus fumigatus  
 US-10-128-714-5563

Query Match 72.4%; Score 15.2; DB 9; Length 4048;  
 Best Local Similarity 85.0%; Pred. No. 2.7e+02;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGATTGCGAGATTGCG 20  
 DB 1016 TCGGGGATTGACACATTGCG 997

## RESULT 14

US-10-270-333-67/c  
 Sequence 67, Application US/10270333  
 Publication No. US20030092124A1  
 GENERAL INFORMATION:  
 APPLICANT: Cravchik, Anibal  
 TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,  
 TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS, AND USES  
 FILE REFERENCE: C1000733CON  
 CURRENT APPLICATION NUMBER: US/10/270,333  
 CURRENT FILING DATE: 2002-10-15  
 PRIOR APPLICATION NUMBER: 60/168,677

PRIOR FILING DATE: 1999-12-03  
 PRIOR APPLICATION NUMBER: 60/175,691  
 PRIOR FILING DATE: 2000-01-12  
 PRIOR APPLICATION NUMBER: 60/191,638  
 PRIOR FILING DATE: 2000-03-23  
 NUMBER OF SEQ ID NOS: 198  
 SOFTWARE: FastSeq for Windows version 4.0  
 SEQ ID NO: 67  
 LENGTH: 9249  
 TYPE: DNA  
 ORGANISM: Drosophila  
 US-10-270-333-67

Query Match 72.4%; Score 15.2; DB 9; Length 9249;  
 Best Local Similarity 85.0%; Pred. No. 2.9e+02;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGATTGCGAGATTGCG 20  
 DB 346 TCGGGGATTGCGAGATTGCG 327

RESULT 15  
 US-10-224-413-3  
 Sequence 3, Application US/10224413  
 Publication No. US20030013167A1  
 GENERAL INFORMATION:

APPLICANT: Ye, Jane et al.  
 TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC  
 TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES  
 FILE REFERENCE: C1001169  
 CURRENT APPLICATION NUMBER: US/10/224,413  
 CURRENT FILING DATE: 2002-08-21  
 PRIOR APPLICATION NUMBER: 09/810,347  
 PRIOR FILING DATE: 2001-03-19  
 NUMBER OF SEQ ID NOS: 6  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO: 3  
 LENGTH: 42571  
 TYPE: DNA  
 ORGANISM: Human  
 US-10-224-413-3

Query Match 72.4%; Score 15.2; DB 9; Length 42571;  
 Best Local Similarity 85.0%; Pred. No. 3.2e+02;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGGGATTGCGAGATTGCGA 21  
 DB 22240 AGGGTAATCTGAGATTGCGA 22259

Search completed: July 8, 2003, 06:03:54  
 Job time : 30.6338 secs

APPLICANT: Hoffman, Neil  
APPLICANT: Hurlan, Patrick  
TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
FILE REFERENCE: 2027 (PARA-016PRV)  
CURRENT APPLICATION NUMBER: US/09/770,444  
CURRENT FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: 60/198,502  
PRIOR FILING DATE: 2000-01-27  
NUMBER OF SEQ ID NOS: 999  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 192  
LENGTH: 470  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-770-444-192

Query Match 72.4%; Score 15.2; DB 10; Length 470;  
Best Local Similarity 85.0%; Pred. No. 2.4e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGATTCGGAGATTGCG 20  
DB 367 TATCGATTCGAGATGCG 348

RESULT 10  
US-09-917-800A-861  
Sequence 881, Application US/09917800A  
Patent No. US20020119462A1  
GENERAL INFORMATION:  
APPLICANT: Mendrick, Donna  
APPLICANT: Porter, Mark  
APPLICANT: Johnson, Kory  
APPLICANT: Castle, Arthur  
APPLICANT: Elashoff, Michael  
TITLE OF INVENTION: Molecular Toxicology Modeling  
FILE REFERENCE: 44921-5038-US  
CURRENT APPLICATION NUMBER: US/09/917,800A  
CURRENT FILING DATE: 2001-07-31  
PRIOR APPLICATION NUMBER: US 60/222,040  
PRIOR FILING DATE: 2000-07-31  
PRIOR APPLICATION NUMBER: US 60/222,880  
PRIOR FILING DATE: 2000-11-02  
PRIOR APPLICATION NUMBER: US 60/290,029  
PRIOR FILING DATE: 2001-05-11  
PRIOR APPLICATION NUMBER: US 60/290,645  
PRIOR FILING DATE: 2001-05-15  
PRIOR APPLICATION NUMBER: US 60/292,336  
PRIOR FILING DATE: 2001-05-22  
PRIOR APPLICATION NUMBER: US 60/295,798  
PRIOR FILING DATE: 2001-06-06  
PRIOR APPLICATION NUMBER: US 60/297,457  
PRIOR FILING DATE: 2001-06-13  
PRIOR APPLICATION NUMBER: US 60/298,884  
PRIOR FILING DATE: 2001-06-19  
PRIOR APPLICATION NUMBER: US 60/303,459  
PRIOR FILING DATE: 2001-07-09  
NUMBER OF SEQ ID NOS: 1740  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 881  
LENGTH: 721  
TYPE: DNA  
ORGANISM: Rattus norvegicus  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20020119462A1 A1170752  
US-09-917-800A-861

Query Match 72.4%; Score 15.2; DB 10; Length 721;  
Best Local Similarity 85.0%; Pred. No. 2.5e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGGATTCGAGATTGCGA 21  
DB 297 AGGGCTTCTGAGACTGCA 316

RESULT 11  
US-09-738-626-595/c  
Sequence 595, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 595  
LENGTH: 1074  
TYPE: DNA  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-595

Query Match 72.4%; Score 15.2; DB 9; Length 1074;  
Best Local Similarity 85.0%; Pred. No. 2.5e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGATTCGGAGATTGCG 20  
DB 568 TAGGGACTCGAGATTGCG 549

RESULT 12  
US-10-128-714-563/c  
Sequence 563, Application US/10128714  
Publication No. US20030119013A1  
GENERAL INFORMATION:  
APPLICANT: Jiang, Bo  
APPLICANT: Hu, Wend  
APPLICANT: Tishkoff, Daniel  
APPLICANT: Zamudio, Carlos  
APPLICANT: Exoskin, Alexey M  
APPLICANT: Lemieux, Sebastien M  
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
FILE REFERENCE: 10182-018-999  
CURRENT APPLICATION NUMBER: US/10/128,714  
CURRENT FILING DATE: 2002-04-23  
PRIOR APPLICATION NUMBER: US 60/285,697  
PRIOR FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: US 60/287,066  
PRIOR FILING DATE: 2001-04-27  
PRIOR APPLICATION NUMBER: US 60/295,890  
PRIOR FILING DATE: 2001-06-05  
PRIOR APPLICATION NUMBER: US 60/303,899  
PRIOR FILING DATE: 2001-07-09  
PRIOR APPLICATION NUMBER: US 60/316,362  
PRIOR FILING DATE: 2001-08-31

```

Db      269 AGGGAATCGGAGATCGCGA 250

RESULT 8
US-09-770-791-820
: Sequence 820, Application US/09770791
: Patent No. US20020062014A1
: GENERAL INFORMATION:
: APPLICANT: Goriach, Jori
: APPLICANT: An, Yong-Qiang
: APPLICANT: Hamilton, Carol M.
: APPLICANT: Price, Jennifer L.
: APPLICANT: Raines, Tracy M.
: APPLICANT: Yu, Yang
: APPLICANT: Rameaka, Joshua G.
: APPLICANT: Page, Amy
: APPLICANT: Mathew, Abraham V.
: APPLICANT: Ledford, Brooke L.
: APPLICANT: Moessner, Jeffrey P.
: APPLICANT: Haas, William David
: APPLICANT: Garcia, Carlos A.
: APPLICANT: Krieker, Maja
: APPLICANT: Slader, Ted
: APPLICANT: Davis, Keith R.
: APPLICANT: Allen, Keith
: APPLICANT: Hoffman, Neil
: APPLICANT: Hurban, Patrick
: TITLE OF INVENTION: Expressed Sequences of Arabidopsis
: TITLE OF INVENTION: thaliana
: FILE REFERENCE: 2029 (PARA-018PRV)
: CURRENT APPLICATION NUMBER: US/09/770,791
: PRIOR FILING DATE: 2001-01-26
: PRIOR APPLICATION NUMBER: 60/178,480
: PRIOR FILING DATE: 2000-01-27
: NUMBER OF SEQ ID NOS: 999
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 820
: LENGTH: 342
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
US-09-770-791-820

Query Match      72.4%; Score 15.2; DB 10; Length 342;
Best Local Similarity 85.0%; Pred. No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0

QY      2 AGGGATTGCGAGATTGCGA 21
      ||||| ||||| ||||| ||
Db      169 AGGGAATCGGAGATAGAGA 188

RESULT 9
US-09-770-444-192/C
: Sequence 192, Application US/09770444
: Patent No. US20020023280A1
: GENERAL INFORMATION:
: APPLICANT: Goriach, Jori
: APPLICANT: An, Yong-Qiang
: APPLICANT: Hamilton, Carol M.
: APPLICANT: Price, Jennifer L.
: APPLICANT: Raines, Tracy M.
: APPLICANT: Yu, Yang
: APPLICANT: Rameaka, Joshua G.
: APPLICANT: Page, Amy
: APPLICANT: Mathew, Abraham V.
: APPLICANT: Ledford, Brooke L.
: APPLICANT: Moessner, Jeffrey P.
: APPLICANT: Haas, William David
: APPLICANT: Garcia, Carlos A.
: APPLICANT: Krieker, Maja
: APPLICANT: Slader, Ted
: APPLICANT: Davis, Keith R.
: APPLICANT: Allen, Keith

```

```
Sequence 29, Application US/09873880
Patent No. US2002012318A1
GENERAL INFORMATION:
APPLICANT: Sewalt, Vincent
APPLICANT: Falco, S. Carl
APPLICANT: Allen, Stephen M.
TITLE OF INVENTION: GLYCINE METABOLISM ENZYMES
FILE REFERENCE: B01192 US CIP
CURRENT APPLICATION NUMBER: US/09/873,880
CURRENT FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 09/363,321
PRIOR FILING DATE: July 28, 1999
PRIOR APPLICATION NUMBER: 60/094,839
PRIOR FILING DATE: July 31, 1998
NUMBER OF SEQ ID NOS: 42
SOFTWARE: Microsoft Office 97
SEQ ID NO 29
LENGTH: 1362
TYPE: DNA
ORGANISM: Glycine max
US-09-873-880-29

Query Match          75.2%, Score 15.8; DB 10; Length 1362;
Best Local Similarity 89.5%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 AGGGATTGCGAGATTGCG 20
DB      360 AGGGATTGCGAGATTGCG 342

RESULT 3
US-10-037-270-446
Sequence 446, Application US/10037270
Publication No. US20030104529A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyun
APPLICANT: Chen, Ruihong
APPLICANT: Zhao, Qing A.
APPLICANT: Weinman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yunqing
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: Tillinshast, John
APPLICANT: Dirmac, Radoje T.
TITLE OF INVENTION: No. US20030104529A1 Nucleic acids and
FILE REFERENCE: Polypeptides
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/10/037,270
CURRENT FILING DATE: 2002-01-04
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1104
SOFTWARE: PC_FL_genes Version 1.0
SEQ ID NO 446
LENGTH: 1471
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE: CDS
NAME/KEY: (904)..(1293)
US-10-037-270-446
```

```
Query Match          73.3%; Score 15.4; DB 9; Length 1471;
Best Local Similarity 94.1%; Pred. No. 2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 GGGATTGCGAGATTGCG 19
DB      669 GAGATTGCGAGATTGCG 685

RESULT 4
US-09-961-527A-6/c
Sequence 6, Application US/09961527A
Patent No. US20020142324A1
GENERAL INFORMATION:
APPLICANT: Wang, Xun
APPLICANT: Turgeon, B. Gillian
APPLICANT: Yoder, Olen
APPLICANT: Wu, Jianguo
TITLE OF INVENTION: Fungal target genes and methods to identify those genes
FILE REFERENCE: TM0129-UT
CURRENT APPLICATION NUMBER: US/09/961,527A
CURRENT FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: US 60/234,673
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: US 60/234,650
PRIOR FILING DATE: 2000-09-22
NUMBER OF SEQ ID NOS: 19
SEQ ID NO 6
LENGTH: 8091
TYPE: DNA
ORGANISM: Cochliobolus
US-09-961-527A-6

Query Match          73.3%; Score 15.4; DB 10; Length 8091;
Best Local Similarity 94.1%; Pred. No. 2.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 GGGATTGCGAGATTGCG 20
DB      5710 GGGATTGCGAGATTGCG 5694

RESULT 5
US-09-961-527A-1/c
Sequence 1, Application US/09961527A
Patent No. US20020142324A1
GENERAL INFORMATION:
APPLICANT: Wang, Xun
APPLICANT: Turgeon, B. Gillian
APPLICANT: Yoder, Olen
APPLICANT: Wu, Jianguo
TITLE OF INVENTION: Fungal target genes and methods to identify those genes
FILE REFERENCE: TM0129-UT
CURRENT APPLICATION NUMBER: US/09/961,527A
CURRENT FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: US 60/234,673
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: US 60/234,650
PRIOR FILING DATE: 2000-09-22
NUMBER OF SEQ ID NOS: 19
SEQ ID NO 1
LENGTH: 14955
TYPE: DNA
ORGANISM: Cochliobolus
US-09-961-527A-1

Query Match          73.3%; Score 15.4; DB 10; Length 14955;
Best Local Similarity 94.1%; Pred. No. 2.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 GGGATTGCGAGATTGCG 20
DB      12555 GGGATTGCGAGATTGCG 12539
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:57:09 ; Search time 10.6338 Seconds  
(without alignments)  
3082.690 Million cell updates/sec

Title: US-09-477-082-29

Perfect score: 21  
1 tagggagattcgagattcgga 21

Sequence: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Scoring table: 1085931 seqs, 780495707 residues

Searched: Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15.8	75.2	495	US-09-873-880-7	Sequence 7, Appl
2	15.8	75.2	1362	US-09-873-880-29	Sequence 29, Appl
3	15.4	73.3	1471	US-10-037-270-446	Sequence 446, Appl
4	15.4	73.3	8091	US-09-861-527A-6	Sequence 6, Appl
5	15.4	73.3	14955	US-09-861-527A-1	Sequence 1, Appl
6	15.2	72.4	282	US-09-294-0938-1232	Sequence 1232, Ap
7	15.2	72.4	285	US-09-294-0938-775	Sequence 775, Ap
8	15.2	72.4	342	US-09-770-791-820	Sequence 820, Ap
9	15.2	72.4	470	US-09-770-444-192	Sequence 192, Ap
10	15.2	72.4	721	US-09-917-800A-881	Sequence 881, Ap
11	15.2	72.4	1074	US-09-738-626-595	Sequence 595, Ap
12	15.2	72.4	4048	US-10-128-714-563	Sequence 563, Ap
13	15.2	72.4	4048	US-10-128-714-563	Sequence 563, Ap
14	15.2	72.4	9249	US-10-224-413-3	Sequence 67, Appl
15	15.2	72.4	42571	US-10-224-413-3	Sequence 3, Appl
16	15.2	72.4	3309400	US-09-738-626-1	Sequence 1, Appl
17	14.6	70.5	387	US-09-878-574-1714	Sequence 1714, Ap
18	14.6	70.5	461	US-09-918-995-35322	Sequence 35322, A
19	14.6	69.5	419	US-10-056-359-1	Sequence 1, Appl

C 20	14.6	69.5	419	12	US-10-056-359-3	Sequence 3, Appl
C 21	14.6	69.5	419	12	US-10-056-360-1	Sequence 1, Appl
C 22	14.6	69.5	419	12	US-10-056-360-3	Sequence 3, Appl
C 23	14.6	69.5	436	10	US-09-878-574-4901	Sequence 4901, Ap
C 24	14.6	69.5	465	10	US-09-960-382-9325	Sequence 9325, Ap
C 25	14.6	69.5	485	9	US-09-770-961-807	Sequence 807, Ap
C 26	14.6	69.5	493	9	US-09-770-961-604	Sequence 604, Ap
C 27	14.6	69.5	1070	9	US-10-205-534-6	Sequence 6, Appl
C 28	14.6	69.5	1206	9	US-09-938-842A-2053	Sequence 2053, Ap
C 29	14.6	69.5	1210	10	US-09-887-576-477	Sequence 477, Ap
C 30	14.6	69.5	1210	10	US-09-887-576-577	Sequence 577, Ap
C 31	14.6	69.5	1302	10	US-09-815-242-7258	Sequence 7258, Ap
C 32	14.6	69.5	1440	10	US-09-881-752A-273	Sequence 273, Ap
C 33	14.6	69.5	2000	10	US-09-887-576-330	Sequence 330, Ap
C 34	14.6	69.5	2004	10	US-09-887-576-174	Sequence 174, Ap
C 35	14.6	69.5	5118	9	US-10-172-086-42	Sequence 42, Appl
C 36	14.6	69.5	6112	9	US-10-172-086-33	Sequence 33, Appl
C 37	14.6	69.5	6284	10	US-09-764-877-3643	Sequence 3643, Ap
C 38	14.6	69.5	7317	9	US-10-239-676-48	Sequence 48, Appl
C 39	14.4	68.6	479	10	US-09-864-761-14337	Sequence 14337, A
C 40	14.4	68.6	509	9	US-09-918-995-7247	Sequence 7247, Ap
C 41	14.4	68.6	972	9	US-10-128-714-7265	Sequence 7265, Ap
C 42	14.4	68.6	981	10	US-09-815-242-7388	Sequence 7388, Ap
C 43	14.4	68.6	1207	9	US-10-128-714-6265	Sequence 6265, Ap
C 44	14.4	68.6	2596	9	US-10-233-131-7	Sequence 7, Appl
C 45	14.4	68.6	2596	10	US-09-808-701-7	Sequence 7, Appl

## ALIGNMENTS

RESULT 1  
US-09-873-880-7/c  
; Sequence 7, Application US/09873880  
; Patent No. US20020123118A1  
; GENERAL INFORMATION:  
; APPLICANT: Sewalt, Vincent  
; APPLICANT: Falco, S. Carl  
; APPLICANT: Allen, Stephen M.  
; TITLE OF INVENTION: GLYCINE METABOLISM ENZYMES  
; FILE REFERENCE: BBI192 US CIP  
; CURRENT APPLICATION NUMBER: US/09/873,880  
; PRIORITY FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: 09/363,321  
; PRIOR FILING DATE: July 28, 1999  
; PRIOR APPLICATION NUMBER: 60/094,839  
; PRIOR FILING DATE: July 31, 1998  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 7  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (382)  
; OTHER INFORMATION: n=A, C, G, or T  
; NAME/KEY: unsure  
; LOCATION: (454)  
; OTHER INFORMATION: n=A, C, G, or T  
US-09-873-880-7

Query Match 75.2%; Score 15.8; DB 10;  
Best Local Similarity 89.5%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGCGATTGCGAGATTGCG 20  
||||||| ||| |||||  
Db 353 AGCGATTGCGAGATTGCG 335

RESULT 2  
US-09-873-880-29/c



OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/00408  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,492  
FILING DATE: 13-January-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: YANKWICH, Leon R  
REGISTRATION NUMBER: 30,237  
REFERENCE/DOCKET NUMBER: TCS-203-PCT(94,664-A)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-345-9100  
TELEFAX: 617-345-9111  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 276 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL:  
ANTI-SENSE:  
FEATURE:  
NAME/KEY:  
LOCATION:  
PCT-US95-00408-25

Query Match 67.6%; Score 14.2; DB 5; Length 276;  
Best Local Similarity 84.2%; Pred. No. 1.8e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
OY 2 AGGGGATTCGAGATTGCG 20  
||| || |||||  
Db 26 AGGTGAGTCGAGATTGCG 8

Search completed: July 8, 2003, 05:56:22  
Job time : 79.0371 secs

sequence 26, Application US/08181492B

COMPUTER: IBM PC Compatible

EARLIER FILING DATE: 1995-07-19  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO: 3  
LENGTH: 419  
TYPE: DNA  
ORGANISM: Papilloma virus  
US-09-000-266-3

Query Match 69.5%; Score 14.6; DB 4; Length 419;  
Best Local Similarity 81.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TAGGGATTCGAGATTGCA 21  
11 |||||||  
DB 390 TATAGATTCGAGCTTCCA 370

RESULT 10  
US-09-628-099-1  
Sequence 1, Application US/09628099  
Patent No. 6368632  
GENERAL INFORMATION:  
APPLICANT: Shamanin, Vladimir  
APPLICANT: De Villiers-Zur Hausen, Ethel-Michele  
TITLE OF INVENTION: PAPILLOMA VIRUSES, AGENTS FOR DETECTING  
FILE REFERENCE: 8484-0037-999  
CURRENT APPLICATION NUMBER: US/09/628,099  
CURRENT FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 09/000,266  
PRIOR FILING DATE: 1998-10-19  
PRIOR APPLICATION NUMBER: DE P 195 26 386.3  
PRIOR FILING DATE: 1995-07-19  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO: 1  
LENGTH: 419  
TYPE: DNA  
ORGANISM: Papilloma virus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(417)  
US-09-628-099-1

Query Match 69.5%; Score 14.6; DB 4; Length 419;  
Best Local Similarity 81.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TAGGGATTCGAGATTGCA 21  
11 |||||||  
DB 30 TATAGATTCGAGCTTCCA 50

RESULT 11  
US-09-628-099-3/C  
Sequence 3, Application US/09628099  
Patent No. 6368852  
GENERAL INFORMATION:  
APPLICANT: Shamanin, Vladimir  
APPLICANT: De Villiers-Zur Hausen, Ethel-Michele  
TITLE OF INVENTION: PAPILLOMA VIRUSES, AGENTS FOR DETECTING  
FILE REFERENCE: 8484-0037-999  
CURRENT APPLICATION NUMBER: US/09/628,099  
CURRENT FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 09/000,266  
PRIOR FILING DATE: 1998-10-19  
PRIOR APPLICATION NUMBER: DE P 195 26 386.3  
PRIOR FILING DATE: 1995-07-19  
NUMBER OF SEQ ID NOS: 30

SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO: 3  
LENGTH: 419  
TYPE: DNA  
ORGANISM: Papilloma virus  
US-09-628-099-3

Query Match 69.5%; Score 14.6; DB 4; Length 419;  
Best Local Similarity 81.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TAGGGATTCGAGATTGCA 21  
11 |||||||  
DB 390 TATAGATTCGAGCTTCCA 370

RESULT 12  
US-09-221-017B-862/c  
Sequence 862, Application US/09221017B  
Patent No. 6444799  
GENERAL INFORMATION:  
APPLICANT: Ross, Bruce C.  
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
NUMBER OF SEQUENCES: 1120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221,017B  
FILING DATE: 23-DEC-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: P1182  
FILING DATE: 31-DEC-1997  
PRIOR APPLICATION DATA: P1546  
FILING DATE: 30-JAN-1998  
PRIOR APPLICATION DATA: P2911  
FILING DATE: 09-APR-1998  
PRIOR APPLICATION DATA: PCT/AU98/01023  
FILING DATE: 10-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Monroy, Gladys H  
REGISTRATION NUMBER: 32,430  
REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 862:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3363 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORYPHYROMONAS GINGIVALIS  
FEATURE:  
NAME/KEY: misc\_feature

```
FEATURE:
NAME/KEY: CDS
LOCATION: (21:81)...(21403)
NAME/KEY: CDS
LOCATION: (95252)...(95430)
NAME/KEY: CDS
LOCATION: (101753)...(101996)
NAME/KEY: CDS
LOCATION: (110324)...(110439)
NAME/KEY: CDS
LOCATION: (124058)...(124278)
NAME/KEY: CDS
LOCATION: (127009)...(127130)
NAME/KEY: CDS
LOCATION: (128910)...(129139)
US-09-426-290-1

Query Match      72.4%; Score 15.2; DB 4; Length 168575;
Best Local Similarity 85.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      2 AGGGGATTGCGAGATTGCGA 21
Db      29456 AGGGGATTGCGAGATTGCGA 29475

RESULT 6
US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      72.4%; Score 15.2; DB 4; Length 4403765;
Best Local Similarity 85.0%; Pred. No. 64;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1 TAGGGGATTGCGAGATTGCG 20
Db      1931100 TAGAGATTGCGAGATTGCG 1931119

RESULT 7
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
```

```
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match      72.4%; Score 15.2; DB 4; Length 4411529;
Best Local Similarity 85.0%; Pred. No. 64;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1 TAGGGGATTGCGAGATTGCG 20
Db      1940329 TAGAGATTGCGAGATTGCG 1940348

RESULT 8
US-09-000-266-1
Sequence 1, Application US/09000266A
Patent No. 6322795
GENERAL INFORMATION:
APPLICANT: Shamanin, Vladimir
APPLICANT: De Villiers-Zur Hausen, Ethel-Michele
APPLICANT: Zur Hausen, Harald
TITLE OF INVENTION: PAPILLOMA VIRUSES, AGENTS FOR DETECTING
TITLE OF INVENTION: THEM AND FOR TREATING DISEASES CAUSED BY SUCH VIRUSES
FILE REFERENCE: 8484-0037-999
CURRENT APPLICATION NUMBER: US/09/000,266A
CURRENT FILING DATE: 1998-10-19
EARLIER APPLICATION NUMBER: PCT/DE96/01369
EARLIER FILING DATE: 1996-07-19
EARLIER APPLICATION NUMBER: DE P 195 26 386.3
EARLIER FILING DATE: 1995-07-19
NUMBER OF SEQ ID NOS: 30
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 419
TYPE: DNA
ORGANISM: Papilloma virus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(417)
US-09-000-266-1

Query Match      69.5%; Score 14.6; DB 4; Length 419;
Best Local Similarity 81.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 TAGGGGATTGCGAGATTGCGA 21
Db      30 TATAGATTGCGAGATTGCGA 50

RESULT 9
US-09-000-266-3/C
Sequence 3, Application US/09000266A
Patent No. 6322795
GENERAL INFORMATION:
APPLICANT: Shamanin, Vladimir
APPLICANT: De Villiers-Zur Hausen, Ethel-Michele
APPLICANT: Zur Hausen, Harald
TITLE OF INVENTION: PAPILLOMA VIRUSES, AGENTS FOR DETECTING
TITLE OF INVENTION: THEM AND FOR TREATING DISEASES CAUSED BY SUCH VIRUSES
FILE REFERENCE: 8484-0037-999
CURRENT APPLICATION NUMBER: US/09/000,266A
CURRENT FILING DATE: 1998-10-19
EARLIER APPLICATION NUMBER: PCT/DE96/01369
EARLIER FILING DATE: 1996-07-19
EARLIER APPLICATION NUMBER: DE P 195 26 386.3
```

TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ. ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2887 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-983-502-14

Query Match 84.8%; Score 17.8; DB 4; Length 2887;  
Best Local Similarity 90.5%; Pred. No. 3.7;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTCGGAGATTGCGA 21  
DB 221 TAGGGGACTCGAGACTGCGA 241

## RESULT 2

PCT-US96-10521-14  
Sequence 14, Application PC/TUS9610521  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS RECEPTORS  
TITLE OF INVENTION: AND OTHER PROTEINS  
NUMBER OF SEQUENCES: 34  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10521  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 114,615  
FILING DATE: 16-JUL-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 114,986  
FILING DATE: 17-AUG-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 115,319  
FILING DATE: 14-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 116,588  
FILING DATE: 27-DEC-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 117,932  
FILING DATE: 16-APR-1996  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2887 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
PCT-US96-10521-14

Query Match 84.8%; Score 17.8; DB 5; Length 2887;  
Best Local Similarity 90.5%; Pred. No. 3.7;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTCGGAGATTGCGA 21  
DB 221 TAGGGGACTCGAGACTGCGA 241

RESULT 3  
US-09-134-246-6/c  
Sequence 6, Application US/09134246B

Patent No. 6207377  
GENERAL INFORMATION:  
APPLICANT: Wayne, Jay  
TITLE OF INVENTION: Method For Construction Of Thermus-E. coli Shuttle  
TITLE OF INVENTION: Vectors And Identification Of Two Thermus Plasmid  
FILE REFERENCE: Thermus Shuttle Vector  
CURRENT APPLICATION NUMBER: US/09/134,246B  
CURRENT FILING DATE: 1998-08-14  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 6  
LENGTH: 5849  
TYPE: DNA  
ORGANISM: Thermus sp.  
US-09-134-246-6

Query Match 72.4%; Score 15.2; DB 4; Length 5849;  
Best Local Similarity 85.0%; Pred. No. 81;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGGATTCGGAGATTGCGA 21  
DB 4909 AGGAGATTAGGAGATTGAGA 4890

## RESULT 4

US-09-810-347-3  
Sequence 3, Application US/09810347  
Patent No. 6461847  
GENERAL INFORMATION:  
APPLICANT: YE, Jane et al.  
TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC  
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES  
FILE REFERENCE: CLO01169  
CURRENT APPLICATION NUMBER: US/09/810,347  
CURRENT FILING DATE: 2001-03-19  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 42571  
TYPE: DNA  
ORGANISM: Human  
US-09-810-347-3

Query Match 72.4%; Score 15.2; DB 4; Length 42571;  
Best Local Similarity 85.0%; Pred. No. 1e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGGATTCGGAGATTGCGA 21  
DB 22240 AGGTAATCTGAGATTGCGA 22259

## RESULT 5

US-09-426-290-1  
Sequence 1, Application US/09426290  
Patent No. 6410712  
GENERAL INFORMATION:  
APPLICANT: Berling Ran Olafsdottir  
TITLE OF INVENTION: HUMAN NARCOLEPSY GENE  
FILE REFERENCE: 2345.2001-000  
CURRENT APPLICATION NUMBER: US/09/426,290  
CURRENT FILING DATE: 1999-10-25  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 168575  
TYPE: DNA  
ORGANISM: Homo Sapiens

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:45:20 ; Search time 5.03708 Seconds

(without alignments)  
1278.561 Million cell updates/sec

Title: US-09-477-082-29

Sequence: 1 taggggattcgagatgcga 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/1/lna/6A.COMB.seq:\*  
4: /cgn2\_6/prodata/1/lna/6B.COMB.seq:\*  
5: /cgn2\_6/prodata/1/lna/PCTUS.COMB.seq:\*  
6: /cgn2\_6/prodata/1/lna/backfilest1.seq:\*  
  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.8	84.8	2887	4	US-08-983-502-14
2	17.8	84.8	2887	4	PCT-US96-10521-14
3	15.2	72.4	5849	4	US-09-134-245-6
4	15.2	72.4	42571	4	US-09-810-347-3
5	15.2	72.4	168575	4	US-09-426-290-1
6	15.2	72.4	4403765	4	US-09-103-840A-2
7	15.2	72.4	4411529	4	US-09-103-840A-1
8	14.6	69.5	419	4	US-09-000-265-1
9	14.6	69.5	419	4	US-09-000-265-3
10	14.6	69.5	419	4	US-09-028-099-1
11	14.6	69.5	419	4	US-09-028-099-3
12	14.6	69.5	3363	4	US-09-221-017B-862
13	14.2	67.6	276	1	US-08-181-492B-25
14	14.2	67.6	276	1	US-08-181-492B-25
15	14.2	67.6	276	5	PCT-US95-00408-25
16	14.2	67.6	1188	4	US-09-064-693A-17
17	14.2	67.6	1312	4	US-09-193-792-20
18	14.2	67.6	2021	4	US-09-193-792-2
19	14.2	67.6	2284	4	US-09-193-792-1
20	14.2	67.6	4086	4	US-08-313-181-1
21	14.2	67.6	6641	4	US-09-064-693A-25
22	14.2	67.6	7445	4	US-09-178-973B-8
23	14.2	67.6	7445	4	US-09-419-568B-8
24	14.2	67.6	7445	4	US-09-354-243B-8
25	14.2	67.6	24417	2	US-08-846-762-1
26	14.2	67.6	33529	4	US-09-144-085-3
27	14.2	67.6	4403765	4	US-09-103-840A-2

C 28	14.2	67.6	4411529	4	US-09-103-840A-1	Sequence 1, Appl1
C 29	13.8	65.7	276	4	US-08-991-789A-32	Sequence 32, Appl1
C 30	13.8	65.7	276	4	US-09-062-451-32	Sequence 32, Appl1
C 31	13.8	65.7	276	4	US-09-598-326-32	Sequence 32, Appl1
C 32	13.8	65.7	420	1	US-08-470-179-162	Sequence 162, App
C 33	13.8	65.7	460	2	US-08-487-727A-1	Sequence 1, Appl1
C 34	13.8	65.7	1484	2	US-08-185-828A-22	Sequence 22, Appl1
C 35	13.8	65.7	1532	2	US-08-185-828A-11	Sequence 11, Appl1
C 36	13.8	65.7	2288	4	US-09-553-132-1	Sequence 1, Appl1
C 37	13.8	65.7	2288	4	US-09-489-964-2	Sequence 2, Appl1
C 38	13.8	65.7	4137	4	US-08-920-812-22	Sequence 22, Appl1
C 39	13.8	65.7	6914	1	US-08-920-812-22	Sequence 22, Appl1
C 40	13.8	65.7	6914	1	US-08-921-177-22	Sequence 22, Appl1
C 41	13.8	65.7	6914	1	US-08-362-577C-22	Sequence 22, Appl1
C 42	13.8	65.7	6914	2	US-08-920-812-22	Sequence 22, Appl1
C 43	13.8	65.7	8912	4	US-08-469-260A-11	Sequence 11, Appl1
C 44	13.8	65.7	9143	2	US-08-639-857-32	Sequence 32, Appl1
C 45	13.8	65.7	9143	2	US-08-639-857-32	Sequence 32, Appl1

## ALIGNMENTS

RESULT 1  
US-08-983-502-14  
; Sequence 14, Application US/08983502  
; Patent No. 6399327  
; GENERAL INFORMATION:  
; APPLICANT: David WALLACH  
; APPLICANT: Mark P. BOLDIN  
; APPLICANT: Tanya M. GONCHAROV  
; APPLICANT: Yury V. GOLTSSEV  
; TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS RECEPTORS  
; TITLE OF INVENTION: AND OTHER PROTEINS  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Broadway and Neimark  
; STREET: 419 Seventh Street N.W., Ste. 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/983,502  
; FILING DATE: 16-JAN-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/10521  
; FILING DATE: 14-JUN-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: IL 114,615  
; FILING DATE: 16-JUL-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: IL 114,986  
; FILING DATE: 17-AUG-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: IL 115,319  
; FILING DATE: 14-SEP-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: IL 116,588  
; FILING DATE: 27-DEC-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: IL 117,932  
; FILING DATE: 16-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Browdy, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: WALLACH-19  
; TELECOMMUNICATION INFORMATION:

```

source
1.722
/organism="Pristionchus pacificus"
/strain="var. California"
/db_xref="taxon:54126"
/clone_lib="Pristionchus pacificus BAC ends"

BASE COUNT      222 a      197 c      146 g      157 t

ORIGIN

Query Match      78.1%; Score 16.4; DB 17; Length 722;
Best Local Similarity 94.4%; Pred. No. 2.2e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 GGGGATTGGAGATTGCG 20
        |||||||
Db      230 GGGGATTGGAGATTGTG 213

RESULT 14
B0135457/c      936 bp      mRNA      linear      EST 23-APR-2002
LOCUS      NF006E01ECF1004 Elicited cell culture Medicago truncatula cDNA
DEFINITION      B0135457
ACCESSION      B0135457
VERSION      B0135457
KEYWORDS      EST.
SOURCE      barrel medic.
ORGANISM      Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 936)
AUTHORS      Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Imanu, J.T., Weller, J.W. and May, G.D.
TITLE      Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
JOURNAL      Unpublished (2000)
COMMENT      Contact: Dixon RA
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org
Insert Length: 936 Std Error: 0.00
Plate: 006 row: E column: 01
Seq primer: TCACACAGCAAGACGCTATGAC.
FEATURES
Source
1.936
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF006E01EC"
/clone_lib="Elicited cell culture"
/tissue_type="Cell cultures derived from root tissues"
/dev_stage="Cell suspensions were subcultured every 14
days. Cells were induced six days after subculture"
/note="Vector: Lambda Zap; Cells were induced with yeast
cell wall extracts equivalent to 50ug/ml glucose in the
final concentration. Samples were taken at 0.5, 1, 12 and
24 hours after induction. Equal amounts of RNA from each
time point were pooled and used for mRNA isolation."

BASE COUNT      272 a      230 c      160 g      274 t

ORIGIN

Query Match      78.1%; Score 16.4; DB 14; Length 936;
Best Local Similarity 94.4%; Pred. No. 2.3e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 GGGATTGGAGATTGCGA 21
        |||||||
Db      594 GGGATTGGAGATTGAGA 577

```

```

RESULT 15
CNS00VWG      280 bp      DNA      linear      GSS 28-JUN-1999
LOCUS      Arabidopsis thaliana genome survey sequence 17 end of BAC T10B23 of
DEFINITION      T10B23 library from strain Columbia of Arabidopsis thaliana, genomic
survey sequence.
ACCESSION      AL092470
VERSION      AL092470.1 GI:5293624
KEYWORDS      GSS.
SOURCE      Arabidopsis thaliana.
ORGANISM      Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 280)
AUTHORS      Salanoubat, M., Choisme, N., Attiguenave, F., Brotier, P., Wincker, P.,
Samsen, D., Saurin, W., Weissenbach, J. and Quetier, F.
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 280)
TITLE      Genoscope.
JOURNAL      Direct Submission
COMMENT      Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
FEATURES
Source
1.280
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="T10B23"
/clone_lib="T10B23"
/note="end : 17"

BASE COUNT      87 a      39 c      43 g      111 t

ORIGIN

Query Match      77.1%; Score 16.2; DB 17; Length 280;
Best Local Similarity 85.7%; Pred. No. 2.2e+03;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TAGGGATTGGAGATTGCGA 21
        |||||||
Db      26 TAGGGATTAGAGATTGCGA 46

```

Search completed: July 8, 2003, 05:51:17  
Job time : 163.34 secs

QY 3 GGGGATTCGAGATTGCG 20  
 |||||  
 Db 234 GGGGATTCGAGATTGCG 217

RESULT 11  
 BH220787 441 bp DNA linear GSS 08-NOV-2001  
 LOCUS 1006097A02.1EL\_X1 1006 - Rescemu Grid G Zea mays genomic, DNA  
 DEFINITION sequence.  
 BH220787  
 ACCESSION BH220787.1 GI:16815046  
 VERSION  
 KEYWORDS GSS.  
 SOURCE  
 ORGANISM Zea mays.  
 Zea mays.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 441)  
 Walbot,V.  
 REFERENCE 1006097A02.1EL\_X1 1006 - Rescemu Grid G Zea mays genomic, DNA  
 TITLE Maize genomic sequences found using engineered Rescemu transposon  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu  
 Very probable ligation site of ends cut by single endonuclease.  
 Reverse complemented post-ligation sequence from source sequence.  
 Plate: 1006097 Row: 29  
 Class: transposon-tagged.  
 Location/Qualifiers  
 1..441  
 /organism="Zea mays"  
 /cultivar="mixed background W23/A188/B73"  
 /db\_xref="taxon:4577"  
 /clone\_id="1006 - Rescemu Grid G"  
 /tissue\_type="leaf"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: leaf; Vector: Rescemu (engineered from  
 Bluescript backbone); Site\_1: BamHI; Site\_2: BglII;  
 Rescemu is a 4.9 kb, modified maize Mu transposon  
 designed to allow plasmid rescue from total genomic DNA.  
 Mu elements insert preferentially into transcription  
 units. For more information on Rescemu, go to the web  
 site 'www.zmdb.lastate.edu' and follow the links for  
 'Rescemu.' Grid G was grown at Stanford in 2000. DNA was  
 extracted from leaf punches, double digested using BamHI  
 and BglII, and ligated to form circular plasmids. DH10B  
 cells were transformed and then screened on LB plates with  
 ampicillin."

BASE COUNT 67 a 162 c 143 g 69 t  
 ORIGIN

Query Match 78.1%; Score 16.4; DB 17; Length 441;  
 Best Local Similarity 94.4%; Pred. No. 2e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GGGGATTCGAGATTGCG 20  
 |||||  
 Db 80 GGGGATTCGAGATTGCG 97

RESULT 12  
 AA390970 598 bp mRNA linear EST 19-APR-2001  
 LOCUS LD09736.5prtime LD Drosophila melanogaster embryo Bluescript  
 DEFINITION Drosophila melanogaster cDNA LD09736 5prtime, mRNA sequence.  
 ACCESSION AA390970  
 VERSION AA390970.1 GI:2043964

KEYWORDS EST.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 598)  
 Harvey,D., Broksstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,  
 Lewis,S. and Rubin,G.M.  
 BDGP/HHMI Drosophila EST project  
 Unpublished (2001)  
 CONTACT: Stapleton, M.  
 BDGP  
 Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798  
 Email: [http://www.fruitfly.org/EST\\_estfruitfly.berkeley.edu](http://www.fruitfly.org/EST_estfruitfly.berkeley.edu)  
 Plate: 97 Row: C Column: 12  
 High quality sequence stop: 469.  
 Location/Qualifiers  
 1..598  
 /organism="Drosophila melanogaster"  
 /db\_xref="BDGP\_EST:BDG1009023"  
 /db\_xref="taxon:7227"  
 /clone="LD09736"  
 /clone\_id="LD Drosophila melanogaster embryo Bluescript"  
 /sex="male and female"  
 /dev\_stage="0 to 24 hours mixed stage embryonic"  
 /lab\_host="SOLR"  
 /note="Organ: embryo; Vector: Bluescript SK; Site\_1: EcoRI  
 ; Site\_2: XhoI; Constructed using StrataGene ZAP-cDNA  
 synthesis kit. Oligo dt-primed and directionally cloned at  
 EcoRI and XhoI in Bluescript SK(+/-)"

BASE COUNT 199 a 105 c 135 g 159 t  
 ORIGIN

Query Match 78.1%; Score 16.4; DB 9; Length 598;  
 Best Local Similarity 94.4%; Pred. No. 2.1e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAGGGATTCGAGATTG 18  
 |||||  
 Db 572 TAGGGATTCGAGATTG 589

RESULT 13  
 BH823627 722 bp DNA linear GSS 20-MAY-2002  
 LOCUS BACP2-006.Y Pristionchus pacificus BAC ends Pristionchus pacificus  
 DEFINITION genomic DNA sequence.  
 BH823627  
 ACCESSION BH823627.1 GI:21001370  
 VERSION  
 KEYWORDS GSS.  
 SOURCE Pristionchus pacificus.  
 ORGANISM Pristionchus pacificus.  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
 Neodiplogasterida; Pristionchus.  
 1 (bases 1 to 722)  
 Sriivasan,J., Sinz,W., Lanz,C., Brand,A., Nandakumar,R., Raddatz  
 G., Witte,H., Keller,H., Kipping,I., Pires da Silva,A., Jesse,T.,  
 Millare,J., de Boich,M., Schuster,S.C. and Sommer,R.J.  
 A BAC-based genetic linkage map of the nematode Pristionchus  
 pacificus  
 Unpublished (2002)  
 CONTACT: Sommer RJ  
 Evolutionary Biology  
 Max-Planck-Institute for Developmental Biology  
 Spemannstr. 37-39, Tuebingen D-72076, Germany  
 Tel: 00497071601371  
 Fax: 00497071601498  
 Email: [raif.sommer@tuebingen.mpg.de](mailto:raif.sommer@tuebingen.mpg.de)  
 Class: BAC ends.  
 Location/Qualifiers



```

/sex="male"
/tissue_type="aorta and vein"
/deg_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGAGAGAGATCCAGACGCTCTTTTTCCTTTTTCVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGATTCGCGATTATTAATAATCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from LambdaBla
FLC T."
BASE COUNT      68 a      91 c      41 g      97 t
ORIGIN
Query Match     78.1%, Score 16.4; DB 10; Length 297;
Best Local Similarity 94.4%; Pred.No.1.8e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY              1 TAGGGGATTCGAGATTG 18
                ||||||| ||||| ||||| |||||
Db              96 TAGGGGTTCGAGATTG 79
RESULT 10
BH826660/c      377 bp DNA linear GSS 20-MAY-2002
LOCUS           BH826660
DEFINITION     BACP24-423.2 Pristionchus pacificus BAC ends Pristionchus
ACCESSION       BH826660
VERSION         BH826660.1 GI:21006923
KEYWORDS        GSS.
SOURCE          Pristionchus pacificus.
ORGANISM        Pristionchus pacificus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE       1 (bases 1 to 377)
AUTHORS        Srinivasan,T., Sinz,W., Lanz,C., Brand,A., Nandakumar,R., Radatz,
               G., Mite,H., Keller,H., Kipping,I., Pires da Silva,A., Jesse,T.,
               Millare,D., de Both,M., Schuster,S.C. and Sommer,R.J.
TITLE           A BAC-based genetic linkage map of the nematode Pristionchus
               pacificus
JOURNAL         Unpublished (2002)
COMMENT         Contact: Sommer RJ
               Evolutionary Biology
               Max-Planck-Institute for Developmental Biology
               Spemannstr. 37-39 Tuebingen D-72076, Germany
               Tel.: 00497071601371
               Fax: 00497071601498
               Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES
   source       1..377
                 /organism="Pristionchus pacificus"
                 /strain="var. Californica"
                 /db_xref="taxon:54126"
                 /clone_lib="Pristionchus pacificus BAC ends"
BASE COUNT     99 a      129 c      66 g      83 t
ORIGIN
Query Match     78.1%, Score 16.4; DB 17; Length 377;
Best Local Similarity 94.4%; Pred.No.1.9e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```



Db 511 TAGGGATTCTGAGATTGC 529

RESULT 4  
BQ226975/c 935 bp mRNA linear EST 02-MAY-2002

LOCUS AGENCOURT\_7561218 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6054859  
5', mRNA sequence.

ACCESSION BQ226975  
VERSION BQ226975.1 GI:20408375  
KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 935)

NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)

UNPUBLISHED (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC/DCMD/DMP  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

Plate: LLM13314 row: c column: 20  
High quality sequence stop: 512.

Location/Qualifiers

1..935

FEATURES

source

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:6054859"

/clone\_lib="NIH\_MGC\_72"

/tissue\_type="melanotic melanoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: skin; Vector: PCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: oligo dt.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

BASE COUNT 225 a 278 c 273 g 159 t

ORIGIN

Query Match 80.0%; Score 16.8; DB 14; Length 935;

Best Local Similarity 90.0%; Pred. No. 1.5e+03;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGGGATTGCGAGATTGCGA 21

Db 797 AGGGGATTCGCGAGATTGCGA 778

RESULT 5

LOCUS BQ899695

DEFINITION AGENCOURT\_8750911 NIH\_MGC\_130 Mus musculus cDNA clone IMAGE:6336113

5', mRNA sequence.

ACCESSION BQ899695

VERSION BQ899695.1 GI:22291697

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 939)

NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)

UNPUBLISHED (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Mark Macconochie, Ph.D. and Nancy L. Freeman,

Ph.D. Library Preparation: Resgen, Invitrogen Corp

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

Plate: LLM13799 row: j column: 18  
High quality sequence stop: 634.

Location/Qualifiers

1..939

FEATURES

source

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="IMAGE:6336113"

/clone\_lib="NIH\_MGC\_130"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: oocytes; Vector: PCMV-SPORT6.1.ccdB;  
Site\_1: EcoRV; Site\_2: NotI; Cloned unidirectionally.  
Primer: oligo dt. Average insert size 1.95 kb.  
Constructed by Resgen, Invitrogen Corp. Note: this is a  
NIH\_MGC Library."

BASE COUNT 250 a 197 c 281 g 210 t 1 others

ORIGIN

Query Match 80.0%; Score 16.8; DB 14; Length 939;

Best Local Similarity 90.0%; Pred. No. 1.5e+03;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGGGATTGCGAGATTGCGA 21

Db 793 AGGGGATTGCGAGATTGCGA 812

RESULT 6

LOCUS BQ672783/c 999 bp mRNA linear EST 15-JUL-2002

DEFINITION AGENCOURT\_8191724 NIH\_MGC\_102 Homo sapiens cDNA clone IMAGE:6257092

5', mRNA sequence.

ACCESSION BQ672783

VERSION BQ672783.1 GI:21783617

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 999)

NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)

UNPUBLISHED (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

Plate: LLM2410 row: n column: 05  
High quality sequence stop: 564.

Location/Qualifiers

1..999

FEATURES

source

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:6257092"

/clone\_lib="NIH\_MGC\_102"

/tissue\_type="epidermoid carcinoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: salivary gland; Vector: POTB; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dt priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAAGAG(G). Library constructed  
by Ling Hong in the laboratory of Gerald M. Rubin

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES  
Location/Qualifiers

1..421

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZP586E18155"  
/clone\_1lb="686 (synonym: hlcc2)"  
/tissue\_type="human skeletal muscle"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="vector: pT7Blue2; Site\_1: SfiIA; Site\_2: SfiIB;  
CDNA-collection"

BASE COUNT 95 a 89 c 113 g 124 t

ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 421;  
Best Local Similarity 90.5%; Pred. No. 4.4e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TAGGGGATTCGAGATTCGA 21  
||||| ||||| ||||| |||||  
DB 284 TAGGGACTCGAGACTCGCA 304

RESULT 2 522 bp mRNA linear EST 14-NOV-2001  
AL600571  
LOCUS DKFZP313F1234.1 313 (synonym: hlcc2) Homo sapiens CDNA clone  
DEFINITION DKFZP313F1234.5, mRNA sequence.  
ACCESSION AL600571  
VERSION AL600571.1 GI:15164077  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 522)  
AUTHORS Ansoerge,W., Winkner,U., Mewes,W., Well,B. and Wiemann,S.  
TITLE EST (Ansoerge,W., Winkner,U., Mewes,H.W., Well,B. and Wiemann,S.)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Ansoerge W

MIPS  
Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by EMBL (European Molecular Biology Laboratories,  
Heidelberg/Germany) within the CDNA sequencing consortium of the  
German Genome Project.  
No sl sequence available.  
This clone (DKFZP313F1234) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES  
Location/Qualifiers

1..522  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZP313F1234"  
/clone\_1lb="313 (synonym: hlcc2)"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="vector: pT7Blue2; Site\_1: SfiIA; Site\_2: SfiIB;  
CDNA-collection"

BASE COUNT 120 a 116 c 133 g 149 t 4 others

ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 522;  
Best Local Similarity 90.5%; Pred. No. 4.4e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TAGGGATTCGAGATTCGA 21  
||||| ||||| ||||| |||||

DB 303 TAGGGACTCGAGACTCGCA 323

RESULT 3 697 bp DNA linear GSS 20-FEB-2001  
A2821487  
LOCUS 2M0094H14F Mouse 10kb plasmid UGCG1M library Mus musculus genomic  
DEFINITION clone UGCG2M0094H14 F, DNA sequence.  
ACCESSION A2821487  
VERSION A2821487.1 GI:12991395  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 697)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,  
M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)

JOURNAL Contact: Robert B. Weiss  
COMMENT University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT  
84112, USA

Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0094 row: H column: 14  
Seq primer: CGTTGTAACGACGCGCCAGT  
Class: plasmid ends  
High quality sequence stop: 697.

FEATURES  
Location/Qualifiers

1..697  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UGCG2M0094H14"  
/clone\_1lb="Mouse 10kb plasmid UGCG1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, TI-resistant, F-"  
/note="Vector: PMD42ny; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PMD42 (g1473214|9b|AF129072.1), a copy-number  
inducible derivative of plasmid RI. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 184 a 189 c 157 g 167 t

ORIGIN

Query Match 82.9%; Score 17.4; DB 17; Length 697;  
Best Local Similarity 94.7%; Pred. No. 7.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAGGGATTCGAGATTCG 19  
||||| ||||| ||||| |||||

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:04:24 : Search time 158.34 Seconds

(without alignments)  
2147.949 Million cell updates/sec

Title: US-09-477-082-29

Perfect score: 21

Sequence: 1 tagggagtcgagatgcga 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues 32308132

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:  
1: em\_estb1:\*  
2: em\_estb2:\*  
3: em\_estb3:\*  
4: em\_estb4:\*  
5: em\_estb5:\*  
6: em\_estb6:\*  
7: em\_estb7:\*  
8: em\_estb8:\*  
9: em\_estb9:\*  
10: em\_estb10:\*  
11: em\_estb11:\*  
12: em\_estb12:\*  
13: em\_estb13:\*  
14: em\_estb14:\*  
15: em\_estb15:\*  
16: em\_estb16:\*  
17: em\_estb17:\*  
18: em\_estb18:\*  
19: em\_estb19:\*  
20: em\_estb20:\*  
21: em\_estb21:\*  
22: em\_estb22:\*  
23: em\_estb23:\*  
24: em\_estb24:\*  
25: em\_estb25:\*  
26: em\_estb26:\*  
27: em\_estb27:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	17.8	84.8	421	9	AL702116	AL702116 DKF2686E
2	17.8	84.8	522	9	AL600571	AL600571 DKF2686E
3	17.4	82.9	697	17	A2821487	A2821487 ZM0094H14
4	16.8	80.0	935	14	BQ226975	BQ226975 AGENCOURT
5	16.8	80.0	939	14	BQ899695	BQ899695 AGENCOURT
6	16.8	80.0	999	14	BQ672783	BQ672783 AGENCOURT

Result No.	Score	Match	Length	DB	ID	Description
7	16.8	80.0	1261	12	BG425308	BG425308 602447220
8	16.4	78.1	233	10	BB171058	BB171058 BB171058
9	16.4	78.1	297	10	BB214821	BB214821 BB214821
10	16.4	78.1	377	17	BB826660	BB826660 BACP24-A
11	16.4	78.1	441	17	BB220787	BB220787 BACP24-A
12	16.4	78.1	598	9	AA390970	AA390970 LD09736.5
13	16.4	78.1	722	17	BB823627	BB823627 BACP2-G0
14	16.4	78.1	936	14	BQ135457	BQ135457 NF0060E1E
15	16.2	77.1	280	17	CNSO0YMG	AL092470 Arabidops
16	16.2	77.1	404	10	AM284750	AM284750 LGL-238A
17	16.2	77.1	466	9	A1947129	A1947129 bs3510.Y
18	16.2	77.1	466	12	BG37945	BG37945 LD09751.3
19	16.2	77.1	511	10	AM078425	AM078425 fe49e03.Y
20	16.2	77.1	519	10	AM058858	AM058858 fe49e03.x
21	16.2	77.1	519	17	DR16H14S	AL48702 Danto rer
22	16.2	77.1	570	13	BM323271	BM323271 PTC1-17D
23	16.2	77.1	613	17	DR18G17S	AL743037 Danto rer
24	16.2	77.1	649	17	A2321056	A2321056 1M0041P02
25	16.2	77.1	673	14	BQ910990	BQ910990 CHA15K03
26	16.2	77.1	738	14	BQ971522	BQ971522 CHA15K03
27	16.2	77.1	874	12	BF260277	BF260277 HVSMEF002
28	16.2	77.1	924	13	BM459635	BM459635 AGENCOURT
29	16.2	77.1	1085	17	AG043322	AG043322 Pan trogl
30	16.2	77.1	1422	17	AG103848	AG103848 Pan trogl
31	16.2	77.1	1502	14	BQ104565	BQ104565 FC0595.e
32	16.2	77.1	1502	17	CNSO16VT	AL107267 Drosophila
33	15.8	75.2	158	10	BE094710	BE094710 CM3-BT077
34	15.8	75.2	177	12	BE845182	BE845182 AD08A09T7
35	15.8	75.2	218	9	AV293343	AV293343 AV293343
36	15.8	75.2	229	10	AM899260	AM899260 CMA-NN008
37	15.8	75.2	324	10	BB102604	BB102604 BB102604
38	15.8	75.2	365	14	BB01102	BB01102 QH14011
39	15.8	75.2	375	12	BF777699	BF777699 NXST-073
40	15.8	75.2	422	10	AV829200	AV829200 AV829200
41	15.8	75.2	470	17	BB194313	BB194313 TC3-70P15
42	15.8	75.2	493	9	AA271049	AA271049 v885c10.r
43	15.8	75.2	493	12	BF056261	BF056261 7K03C07.X
44	15.8	75.2	521	14	BQ39223	BQ39223 MR0-NN019
45	15.8	75.2	533	9	A1774687	A1774687 EST255787

## ALIGNMENTS

RESULT 1  
AL702116  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AL702116 421 bp mRNA linear EST 22-MAR-2002  
DKF2686E18155.1 686 (synonym: hlc03) Homo sapiens cDNA clone  
DKF2686E18155.5, mRNA sequence.  
AL702116.1 GI:19685471  
EST.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Bloecker/H., Boecher/M., Brandt/P., Mewes/W., Well/B. and Wiemann  
/S.  
EST (Bloecker/H., Boecher/M., Brandt/P., Mewes/W., Well/B. and Wiemann  
/S.)  
Unpublished (1999)  
Contact: Bloecker H  
MIPS  
Am Kiofierspitz 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;  
Breuninger/Germany) within the cDNA sequencing consortium of the  
German Genome Project.  
No si sequence available.  
This clone (DKF2686E18155) is available at the RZPD in Berlin.

Query Match 80.0%; Score 16.8; DB 24; length 811;  
 Best Local Similarity 90.0%; Pred. No. 1e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGGGATTTCGGAGATTCGCA 21  
 |||||  
 Db 247 AGGGGATTTCGGAGATTCGAGA 228

Search completed: July 8, 2003, 03:11:11  
 Job time : 21.9513 secs

CC methylation status of many C residues to be determined simultaneously.  
CC ABQ31410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
SQ Sequence 540 BP; 197 A; 200 C; 57 G; 86 T; 0 other;  
Query Match 80.0%; Score 16.8; DB 24; Length 540;  
Best Local Similarity 90.0%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TAGGGGATTCGAGATTGCG 20  
DB 84 TTGGGATTCGAGATTGCG 65  
RESULT 14  
ABQ31406  
ID ABQ31406 standard; DNA; 811 BP.  
AC ABQ31406;  
XX  
DT 12-JUL-2002 (first entry)  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 17997.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP10074.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
PA (EPiG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
DR WPI; 2002-371829/40.  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ31410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in

CC the disclosure of the invention.  
XX  
SQ Sequence 811 BP; 126 A; 97 C; 296 G; 292 T; 0 other;  
Query Match 80.0%; Score 16.8; DB 24; Length 811;  
Best Local Similarity 90.0%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AGGGGATTCGAGATTGCCA 21  
DB 565 AGGGGATTCGAGATCGAGA 584  
RESULT 15  
ABQ31407/C  
ID ABQ31407 standard; DNA; 811 BP.  
AC ABQ31407;  
XX  
DT 12-JUL-2002 (first entry)  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 17998.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP10074.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
PA (EPiG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
DR WPI; 2002-371829/40.  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ31410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
SQ Sequence 811 BP; 292 A; 296 C; 97 G; 126 T; 0 other;

PS Claim 1; SEQ ID NO 767; 32pp + Sequence Listing; German.  
XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/intestinal bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
SQ Sequence 8951 BP; 1965 A; 548 C; 2538 G; 3900 T; 0 other;  
Query Match 82.9%; Score 17.4; DB 24; Length 8951;  
Best Local Similarity 94.7%; Pred. No. 58;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3 GGGGATCGGAGATTGCGA 21  
DB 5581 GGGGTTTCGAGATTGCGA 5599  
RESULT 12  
ABQ22580  
ID ABQ22580 standard; DNA; 540 BP.  
XX  
AC ABQ22580;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 9171.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP10074.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
DR WPI; 2002-371829/40.  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory

CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
SQ Sequence 540 BP; 86 A; 57 C; 200 G; 197 T; 0 other;  
Query Match 80.0%; Score 16.8; DB 24; Length 540;  
Best Local Similarity 90.0%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TAGGGGATCGGAGATTGCG 20  
DB 457 TTGGGATTCGGAGATTGCG 476  
RESULT 13  
ABQ22581/c  
ID ABQ22581 standard; DNA; 540 BP.  
XX  
AC ABQ22581;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 9172.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP10074.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
DR WPI; 2002-371829/40.  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the



PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241122.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244611.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2000US-0259678.  
 XX  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM:  
 XX  
 XX WPI: 2001-483426/52.  
 DR  
 XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and

PT metastasis -  
 XX  
 PS Disclosure; SEQ ID NO 32029; 3071pp + Sequence Listing; English.  
 XX  
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patients own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/haematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 45017 BP; 12919 A; 10037 C; 9566 G; 12495 T; 0 other:  
 Query Match 84.8%; Score 17.8; DB 22; Length 45017;  
 Best Local Similarity 90.5%; Pred. No. 40;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 TAGGGATTGGAGATTGCGA 21  
 Db 25422 TAGGGACTCGAGACTGCGA 25402  
 RESULT 11  
 ID ABL32794 standard; DNA: 8951 BP.  
 XX  
 AC ABL32794;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 767.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antihaemic; cytosolic; nootropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 KW  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200200928-A2.  
 XX  
 PF 03-JAN-2002.  
 XX  
 FE 02-JUL-2001; 2001WO-EP07537.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI: 2002-130909/17.  
 XX  
 XX Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 XX



RESULT 9  
AAK77216/C  
ID AAK77216 standard; DNA; 43338 BP.  
XX  
AC AAK77216;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32028.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX cytostatic; gene therapy; vaccine; metastasis; ds.  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
PD  
XX 09-AUG-2001.  
PF  
XX 17-JAN-2001; 2001WO-US01354.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 02-MAR-2000; 2000US-0184664.  
PR 16-MAR-2000; 2000US-0186350.  
PR 17-MAR-2000; 2000US-0189874.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0217496.  
PR 26-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0225759.  
PR 22-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0227009.  
PR 01-SEP-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0228927.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229512.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0233080.  
PR 08-SEP-2000; 2000US-0233081.  
PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0234984.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 02-OCT-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241121.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246529.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249247.  
PR 17-NOV-2000; 2000US-0249248.  
PR 17-NOV-2000; 2000US-0249249.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.

FT /product= "Human caspase-8"  
 XX WO200129232-A2.  
 XX  
 XX 26-APR-2001.  
 XX  
 XX 19-OCT-2000; 2000WO-US28941.  
 XX  
 XX 20-OCT-1999; 99US-0160559.  
 XX 14-AUG-2000; 2000US-0225564.  
 XX  
 XX (SCIO-) SCIOS INC.  
 XX  
 XX Cordell B, Li Y;  
 XX  
 XX WPI: 2001-290920/30.  
 XX P-PSDB; AAE00605.  
 XX  
 XX Novel fusion polypeptide comprising first and second caspase subunit  
 PT separated by cleavage site not associated in nature with caspase  
 PT subunit, useful for cloning gene encoding enzymes involved in  
 PT proteolytic cleavage -  
 XX  
 XX Disclosure; Fig 15; 116pp; English.  
 XX  
 XX The present sequence is a DNA encoding human Caspase-8 also known as  
 CC MACH, FICE and Mch5. Caspases are a family of cysteine proteases, that  
 CC participate in the initiation and execution of apoptosis. Caspases exist  
 CC as pro-enzymes, activated by cleavage into a large and small subunit,  
 CC occurring after specific aspartic acid residues within the pro-enzyme  
 CC sequence. The present invention relates to a method for functional  
 CC cloning of genes encoding proteins or enzymes involved in proteolytic  
 CC cleavage. The invention is based on the use of caspase expression  
 CC cassettes comprising the coding sequence of a proteolytic cleavage site  
 CC flanked by sequences encoding two caspase subunits. A fusion polypeptide  
 CC comprising a first and a second caspase subunit, separated by a cleavage  
 CC site not associated in nature, is useful for cloning gene encoding  
 CC enzymes involved in proteolytic cleavage. An expression cassette  
 CC containing fusion polypeptide is used to identify a mutant cell line  
 CC deficient in an enzyme of interest and is also useful for diagnosis and  
 CC suppression of proliferation or metastases of a tumour cell characterised  
 CC by overexpression of a polypeptide (e.g. Cathepsin B or urokinase,  
 CC selectively expressed in the tumour cells). DNA encoding fusion  
 CC polypeptide is used in gene therapy.  
 CC  
 XX Sequence 2887 BP; 853 A; 595 C; 625 G; 814 T; 0 other;  
 XX  
 XX Query Match 84.8%; Score 17.8; DB 22; Length 2887;  
 XX Best Local Similarity 90.5%; Pred. No. 36;  
 XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 XX  
 OY 1 TAGGGGATTGGAGATTGCGA 21  
 XX ||||||| ||||||| |||||  
 DB 221 TAGGGGACTCGAGACTGCGA 241  
 XX  
 XX RESULT 8  
 XX ABR84723  
 XX ID ABR84723 standard; cDNA; 2887 BP.  
 XX  
 XX AC ABR84723;  
 XX  
 DT 14-AUG-2002 (first entry)  
 XX  
 XX Human cDNA differentially expressed in granulocytic cells #1294.  
 XX  
 XX Human; ss; granulocytic cell; DNA chip; bacterial infection;  
 KW viral infection; parasitic infection; protozoal infection;  
 KW fungal infection; sterile inflammatory disease; psoriasis;  
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
 KW adult respiratory distress syndrome; inflammatory bowel disease;  
 KW Crohn's disease; ulcerative colitis; periodontal disease;

KW granulocyte activation; chronic inflammation; allergy.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200228999-A2.  
 XX  
 XX 11-APR-2002.  
 XX  
 XX 03-OCT-2001; 2001WO-US30821.  
 XX  
 XX 03-OCT-2000; 2000US-237189P.  
 XX  
 XX (GENE-) GENE LOGIC INC.  
 XX  
 XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
 XX  
 XX WPI: 2002-435328/46.  
 XX  
 XX Detecting granulocyte activation by detecting differential expression  
 PT of genes associated with granulocyte activation, which serves as  
 PT diagnostic markers that is useful for monitoring disease states and  
 PT drug toxicity -  
 XX  
 XX Claim 1; SEQ ID NO 1294; 114pp; English.  
 XX  
 XX The invention relates to detecting (M1) granulocyte (GC) activation  
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
 CC DNA chip analysis as given in the specification, and comparing  
 CC the expression level to an expression level in an unactivated  
 CC GC, where differential expression of Gs is indicative of GCA.  
 CC Also included are modulating (M2) GCA by contacting GC with an agent  
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)  
 CC for an agent capable of modulating GCA or an inflammation (especially  
 CC chronic) in a tissue, an allergic response in a subject, exposure of a  
 CC subject to a pathogen or sterile inflammatory disease using the  
 CC gene expression profile; (3) detecting (M4) an inflammation (especially  
 CC chronic) in a tissue, an allergic response in a subject, exposure of a  
 CC subject to a pathogen or sterile inflammatory disease, by detecting the  
 CC level of expression in a sample of the tissue of gene(s) from Gs, where  
 CC the level of expression of the gene is indicative of inflammation;  
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,  
 CC an allergic response in a subject, exposure of a subject to a pathogen  
 CC or sterile inflammatory disease, by contacting a tissue having  
 CC inflammation with an agent that modulates the expression of gene(s)  
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
 CC modulating GCA; M3 is useful for screening an agent capable of modulating  
 CC GCA preferably in an inflammation in a tissue; M4 is useful for  
 CC detecting an inflammation (especially chronic) in a tissue, an allergic  
 CC response in a subject, exposure of a subject to a pathogen or sterile  
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,  
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal  
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,  
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,  
 CC periodontal disease; also bacterial infection, viral infection,  
 CC parasitic infection, protozoal infection, fungal infection and M5 is  
 CC useful for treating one of the above conditions. The present  
 CC sequence represents a gene differentially expressed in granulocytes.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 2887 BP; 853 A; 595 C; 625 G; 814 T; 0 other;  
 XX  
 XX Query Match 84.8%; Score 17.8; DB 24; Length 2887;  
 XX Best Local Similarity 90.5%; Pred. No. 36;  
 XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 XX  
 OY 1 TAGGGGATTGGAGATTGCGA 21  
 XX ||||||| ||||||| |||||  
 DB 221 TAGGGGACTCGAGACTGCGA 241  
 XX

```

XX Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or
PT prognosing cancer, comprises detecting a modification of genomic DNA
PT comprising the CASP8 gene that results in inactivation of the gene
PS
PS Claim 21, Page 97-98; 107pp; English.
XX
XX Inactivation of a caspase-8 (CASP8) gene is determined by detecting a
CC modification of CASP8 genomic DNA that results in inactivation of the
CC gene. CASP8, a cysteine protease, is part of the death induction signaling
CC complex (DISC), associated with the Fas receptor. CASP8 is inactivated in
CC cancers, and plays a role of a tumor suppressor gene. The CASP8 promoter
CC region sequences, in particular Region 1 and Region 2, are crucial to the
CC design and execution of the genomic methylation PCR analysis of CASP8
CC gene inactivation. Methylation PCR can be used to examine even minute
CC amounts of patient material to demonstrate whether the CASP8 gene
CC expresses an mRNA and protein product. The promoter Region 1 sequence is
CC located upstream (5') to exon 1, which is the alternatively-spliced 5'
CC untranslated region (UTR) that is less commonly used. The promoter
CC Region 2 sequence is located downstream (3') of exon 1 and upstream of
CC exon 2, which is the more commonly used 5' UTR. The CASP8 gene has been
CC localized to human chromosome 2q33-34. The methods are used to diagnose
CC or prognose cancer. Cancer is treated by administering a vector that
CC expresses a gene encoding functional CASP8 in cells. The cancer that
CC is diagnosed or treated is a tumour in which a myc gene is amplified,
CC such as a neuroblastoma. Aggressive neuroblastoma, juvenile neuroblastoma
CC (preferred), small-cell lung carcinoma, non-small-cell lung carcinoma,
CC colorectal carcinoma, or uterine cervical carcinoma can be diagnosed with
CC the new method. A kit for screening for a compound that induces
CC death-receptor-mediated apoptosis in cells containing an inactivated
CC CASP8 gene is also provided.
XX
XX Sequence 753 BP; 140 A; 204 C; 171 G; 238 T; 0 other:
SQ
Query Match      84.8%; Score 17.8; DB 21; Length 753;
Best Local Similarity 90.5%; Pred. No. 34;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY      1 TAGGGGATTGGAGATTGCGA 21
      ||||||| ||||||| |||||
Db      90 TAGGGGACTCGAGACTGCGA 110

```

RESULT 6  
ID AAT61405 standard; cDNA; 2887 BP.  
XX  
XX AAT61405;  
XX  
XX 29-OCT-1997 (first entry)  
XX  
XX MACH isoform alpha coding sequence.  
XX  
XX MACH; MORT-1 binding protein; mediator of receptor toxicity; cell death;  
XX antibody; Fas ligand receptor; Fas-R; death domain region; septic shock;  
XX tumour necrosis factor; HIV-infection; oligodendrocyte death;  
XX apoptosis/programmed cell death; p55-R; graft rejection; acute hepatitis;  
XX autoimmune disease; multiple sclerosis; AIDS-inhibited T-cell suicide;  
XX TNF; therapy; ss.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX CDS 291..1731  
XX FT /\*tag= a  
XX ET /product= MORT-1 binding protein (MACH) isoform alpha  
XX  
XX WO9703998-A1.  
XX  
XX 06-FEB-1997.  
XX  
XX 14-JUN-1996; 96WO-US10521.  
XX

```

PR 16-APR-1996; 96IL-0117932.  
PR 16-JUL-1995; 95IL-0114615.  
PR 17-AUG-1995; 95IL-0114986.  
PR 14-SEP-1995; 95IL-0115319.  
PR 27-SEP-1995; 95IL-0116588.  
XX  
XX (WEIN/) WEINMURZEL H.  
XX (YEDA ) YEDA RES & DEV CO LTD.  
XX  
XX Boldin M, Goltsev YV, Goncharov T, Wallach D;  
XX WPI: 1997-132570/12.  
XX P-PSDB; AAM11891.  
XX  
XX New DNA encoding MACH protein that interacts with MORT-1 protein -  
XX to mediate intracellular effects of Fas or TNF receptors, partic.  
XX for regulating apoptosis in tumours, virus-infected cells etc.  
XX  
XX Example 3; Page 116-117; 163pp; English.  
XX  
XX AAT61396, and AAT61405-T61411 represent coding sequences for different  
XX isoforms of MACH. MACH is a binding protein for the mediator of receptor  
XX toxicity (MORT-1) protein. MORT-1 binds to the Fas ligand receptor  
XX (Fas-R) death domain region, and triggers part of the cell death  
XX signalling cascade in mammalian cells. Vectors containing these  
XX sequences, the encoded proteins, and antibodies (Ab) against them are  
XX used to modulate the effect of Fas-R ligand or TNF on cells that carry  
XX Fas-R or p55-R. This is specifically for treating tumours, HIV-infected  
XX cells or other diseased cells, by control of apoptosis/programmed cell  
XX death. The encoded proteins are mediators of the cell death pathway  
XX initiated by TNF and Fas-R binding, i.e. it mimics or enhances the  
XX effect of MORT-1 where increased cytotoxicity is required. To inhibit the  
XX effect of MORT-1, e.g. in cases of septic shock, graft rejection and  
XX acute hepatitis, sequences encoding antisense molecules or ribozymes, or  
XX Ab against the protein, are used. Compounds that inhibit MACH are  
XX potentially useful for controlling MACH activity e.g. in cases of  
XX autoimmune disease, oligodendrocyte death in multiple sclerosis or  
XX AIDS-inhibited T-cell suicide. The MACH proteins can also be used to  
XX isolate and characterise other proteins and receptors involved in  
XX signalling and for Ab production. The Ab can be used to purify the new  
XX proteins and for diagnosis of conditions involving abnormal function of  
XX Fas-R mediated cellular effects.  
XX  
XX Sequence 2887 BP; 853 A; 595 C; 625 G; 814 T; 0 other:  
SQ
Query Match      84.8%; Score 17.8; DB 18; Length 2887;  
Best Local Similarity 90.5%; Pred. No. 36;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY      1 TAGGGGATTGGAGATTGCGA 21
      ||||||| ||||||| |||||
Db      221 TAGGGGACTCGAGACTGCGA 241

```

RESULT 7  
ID AAD03915 standard; DNA; 2887 BP.  
XX  
XX AAD03915;  
XX  
XX 02-JUL-2001 (first entry)  
XX  
XX Human caspase-8 DNA.  
XX  
XX Human; caspase-8; MACH; FLICE; Mch3;  
XX cysteine protease; apoptosis; caspase expression cassette; metastasis;  
XX tumour; cathepsin B; urokinase; proliferation; gene therapy; ds.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX CDS 292..1731  
XX FT /\*tag= a  
XX

PT Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or  
 PT prognosing cancer, comprises detecting a modification of genomic DNA  
 PT comprising the CASP8 gene that results in inactivation of the gene  
 PS Example 3; Page 61; 107pp; English.

XX Investigation of silencing of caspase-8 (CASP8) gene by methylation was  
 CC analyzed using methylation-sensitive PCR analysis. Amplification of the  
 CC 5' untranslated region (UTR) of the CASP8 gene was performed in reaction  
 CC mixtures containing bisulfite treated DNA. Primers AA51818-23 were  
 CC designed to produce a 330 bp fragment in the upstream region of Casp8  
 CC gene extending from nucleotides +221 to +541. Wild type primers were  
 CC used to amplify the corresponding region of untreated genomic DNA.  
 CC Controls without DNA were also performed. CASP8, a cysteine protease, is  
 CC part of the death inducing signaling complex (DISC) associated with the  
 CC Fas receptor. CASP8 is inactivated in cancers, and plays a role of a  
 CC tumour suppressor gene. The CASP8 promoter region sequences, in  
 CC particular Region 1 and Region 2, are crucial to the design and execution  
 CC of the genomic methylation PCR analysis of CASP8 gene inactivation.  
 CC Methylation PCR can be used to examine even minute amounts of patient  
 CC material to demonstrate whether the CASP8 gene expresses an mRNA and  
 CC protein product. The CASP8 gene has been localized to human chromosome  
 CC 2q33-34. The methods are used to diagnose or prognose cancer. Cancer is  
 CC treated by administering a vector that expresses a gene encoding  
 CC functional CASP8 in cells. The cancer that is diagnosed or treated is a  
 CC tumour in which a myc gene is amplified, such as a neuroblastoma.  
 CC Aggressive neuroblastoma, juvenile neuroblastoma (preferred), small-cell  
 CC lung carcinoma, non-small-cell lung carcinoma, colorectal carcinoma, or  
 CC uterine cervical carcinoma can be diagnosed with the new method. A kit  
 CC for screening for a compound that induces death-receptor-mediated  
 CC apoptosis in cells containing an inactivated CASP8 gene is also  
 CC provided.

XX SQ Sequence 21 BP; 5 A; 4 C; 9 G; 3 T; 0 other;

Query Match 84.8%; Score 17.8; DB 21; Length 21;  
 Best Local Similarity 90.5%; Pred. No. 30;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TAGGGGATTCGGAGATTGCGA 21  
 ||||| ||||| ||||| |||||  
 Db 1 TAGGGGACTCGAGACTGCGA 21

RESULT 4

ABN58902  
 ID ABN58902 standard; DNA; 60 BP.

XX AC ABN58902;

DT 15-JUL-2002 (first entry)

DE Human spliced transcript detection oligonucleotide SEQ ID NO:31650.

KW Human; mouse; rat; splice transcript; detection; RNA transcript;

KM splice variant; transcriptome; oligonucleotide library; ss.

OS Homo sapiens.

WO200210449-A2.

PD 07-FEB-2002.

PF 20-JUL-2001; 2001WO-1B01903.

PR 28-JUL-2000; 2000US-221607P.

PR 02-MAY-2001; 2001US-287724P.

PA (COMP-) COMPUGEN INC.

PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;

XX WPI; 2002-257383/30.

XX New oligonucleotide libraries comprising oligonucleotides which  
 PT selectively hybridize to mRNAs transcribed from a transcription unit of  
 PT a genome, useful for detecting tissue-, pathology-, and  
 PT developmental-specific genes.

PS Example 1; SEQ ID 31650; 47pp; English.

XX The present invention describes oligonucleotide libraries for detecting  
 CC messenger RNAs that populate a (sub-)transcriptome, where the  
 CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple  
 CC transcription units that populate a genome. The library comprises  
 CC several oligonucleotides, each capable of hybridizing selectively to a  
 CC set of messenger RNAs transcribed from a given transcription unit of  
 CC the genome, which encodes one or more messenger RNA splice variants.  
 CC The oligonucleotide libraries are useful for detecting mRNAs from a  
 CC biological sample, in expression profiling studies, in qualitatively or  
 CC quantitatively characterizing the corresponding transcriptome, and in  
 CC detecting RNA transcripts and splice variants of human or animal  
 CC transcriptomes. The libraries may also be used as specialised mini  
 CC libraries to detect transcripts of a sub-transcriptome under a  
 CC particular biological or pathological state, and so allowing the  
 CC detection of tissue- and pathology-specific genes such as those genes  
 CC only expressed in specific tissue under a specific pathological  
 CC condition, to detect developmental specific genes; and to detect RNA  
 CC transcripts and splice variants of a transcriptome of a patient suffering  
 CC from a particular disorder. ABN27253 to ABN59589 represent  
 CC oligonucleotide sequences from rats, humans and mice, which are used in  
 CC the exemplification of the present invention.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 60 BP; 14 A; 8 C; 28 G; 10 T; 0 other;

Query Match 84.8%; Score 17.8; DB 24; Length 60;  
 Best Local Similarity 90.5%; Pred. No. 31;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TAGGGGATTCGGAGATTGCGA 21  
 ||||| ||||| ||||| |||||  
 Db 19 TAGGGGACTCGAGACTGCGA 39

RESULT 5

AA51793  
 ID AA51793 standard; DNA; 753 BP.

XX AC AA51793;

DT 31-OCT-2000 (first entry)

DE CASP8 promoter Region 2.

KW CASP8; caspase-8; promoter; Region 2; inactivation; methylation;

KM Cysteine protease; tumour suppressor; chromosome 2q33-34; neuroblastoma;

KM cancer; death receptor; apoptosis; cytostatic; gene therapy; ss.

OS Homo sapiens.

WO200039347-A1.

PD 06-JUL-2000.

PF 30-DEC-1999; 99WO-US31280.

PR 31-DEC-1998; 98US-0114308.

PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

PI Kidd VJ, Lantti JM, Teltz T;

XX WPI; 2000-452423/39.

```
PT comprising the CASP8 gene that results in inactivation of the gene
XX
PS Example 3; Page 61; 107pp; English.
XX
CC Investigation of silencing of caspase-8 (CASP8) gene by methylation was
CC analyzed using methylation-sensitive PCR analysis. Amplification of the
CC 5' untranslated region (UTR) of the Casp8 gene was performed in reaction
CC mixtures containing bisulfite treated DNA. Primers AA51818-23 were
CC designed to produce a 320 bp fragment in the upstream region of Casp8
CC gene extending from nucleotides +221 to +541. Wild type primers were
CC used to amplify the corresponding region of untreated genomic DNA.
CC Controls without DNA were also performed. CASP8, a cysteine protease, is
CC part of the death inducing signaling complex (DISC) associated with the
CC Fas receptor. CASP8 is inactivated in cancers, and plays a role of a
CC tumour suppressor gene. The CASP8 promoter region sequences, in
CC particular Region 1 and Region 2, are crucial to the design and execution
CC of the genomic methylation PCR analysis of CASP8 gene inactivation.
CC Methylation PCR can be used to examine even minute amounts of patient
CC material to demonstrate whether the CASP8 gene expresses an mRNA and
CC protein product. The CASP8 gene has been localized to human chromosome
CC 2q33-34. The methods are used to diagnose or prognose cancer. Cancer is
CC treated by administering a vector that expresses a gene encoding
CC functional CASP8 in cells. The cancer that is diagnosed or treated is a
CC tumour in which a myc gene is amplified, such as a neuroblastoma.
CC Aggressive neuroblastoma, juvenile neuroblastoma (preferred), small-cell
CC lung carcinoma, non-small-cell lung carcinoma, colorectal carcinoma, or
CC uterine cervical carcinoma can be diagnosed with the new method. A kit
CC for screening for a compound that induces death-receptor-mediated
CC apoptosis in cells containing an inactivated CASP8 gene is also
CC provided.
XX
SQ Sequence 21 BP; 5 A; 2 C; 9 G; 5 T; 0 other;
XX
Query Match 100.0%; Score 21; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAGGGGATTGCGAGATTGCGA 21
Db 1 TAGGGGATTGCGAGATTGCGA 21
RESULT 2
AA51820
ID AA51820 standard; DNA; 21 BP.
XX
AC AA51820;
XX
DT 31-OCT-2000 (first entry)
XX
DE Sense primer for CASP8 unmethylated-specific bisulfite treated DNA.
XX
XX CASP8; caspase-8; promoter; inactivation; methylation; cysteine protease;
XX tumour suppressor; chromosome 2q33-34; neuroblastoma; cancer;
XX death receptor; apoptosis; cytostatic; gene therapy; primer; ss.
XX
OS Homo sapiens.
XX
PN WO200039347-A1.
XX
PD 06-JUL-2000.
XX
PF 30-DEC-1999; 99WO-US31280.
XX
PR 31-DEC-1998; 98US-0114308.
XX
PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
PI Kidd VJ, Lahti JM, Teitz T;
XX
DR WPI; 2000-452423/39.
XX
PT Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or
```

```
PT prognosing cancer, comprises detecting a modification of genomic DNA
XX comprising the CASP8 gene that results in inactivation of the gene
XX
PS Example 3; Page 61; 107pp; English.
XX
CC Investigation of silencing of caspase-8 (CASP8) gene by methylation was
CC analyzed using methylation-sensitive PCR analysis. Amplification of the
CC 5' untranslated region (UTR) of the Casp8 gene was performed in reaction
CC mixtures containing bisulfite treated DNA. Primers AA51818-23 were
CC designed to produce a 320 bp fragment in the upstream region of Casp8
CC gene extending from nucleotides +221 to +541. Wild type primers were
CC used to amplify the corresponding region of untreated genomic DNA.
CC Controls without DNA were also performed. CASP8, a cysteine protease, is
CC part of the death inducing signaling complex (DISC) associated with the
CC Fas receptor. CASP8 is inactivated in cancers, and plays a role of a
CC tumour suppressor gene. The CASP8 promoter region sequences, in
CC particular Region 1 and Region 2, are crucial to the design and execution
CC of the genomic methylation PCR analysis of CASP8 gene inactivation.
CC Methylation PCR can be used to examine even minute amounts of patient
CC material to demonstrate whether the CASP8 gene expresses an mRNA and
CC protein product. The CASP8 gene has been localized to human chromosome
CC 2q33-34. The methods are used to diagnose or prognose cancer. Cancer is
CC treated by administering a vector that expresses a gene encoding
CC functional CASP8 in cells. The cancer that is diagnosed or treated is a
CC tumour in which a myc gene is amplified, such as a neuroblastoma.
CC Aggressive neuroblastoma, juvenile neuroblastoma (preferred), small-cell
CC lung carcinoma, non-small-cell lung carcinoma, colorectal carcinoma, or
CC uterine cervical carcinoma can be diagnosed with the new method. A kit
CC for screening for a compound that induces death-receptor-mediated
CC apoptosis in cells containing an inactivated CASP8 gene is also
CC provided.
XX
SQ Sequence 21 BP; 5 A; 0 C; 9 G; 7 T; 0 other;
XX
Query Match 84.8%; Score 17.8; DB 21; Length 21;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TAGGGGATTGCGAGATTGCGA 21
Db 1 TAGGGGATTGCGAGATTGCGA 21
RESULT 3
AA51822
ID AA51822 standard; DNA; 21 BP.
XX
AC AA51822;
XX
DT 31-OCT-2000 (first entry)
XX
DE Sense primer for untreated CASP8 wild type DNA.
XX
XX CASP8; caspase-8; promoter; inactivation; methylation; cysteine protease;
XX tumour suppressor; chromosome 2q33-34; neuroblastoma; cancer;
XX death receptor; apoptosis; cytostatic; gene therapy; primer; ss.
XX
OS Homo sapiens.
XX
PN WO200039347-A1.
XX
PD 06-JUL-2000.
XX
PF 30-DEC-1999; 99WO-US31280.
XX
PR 31-DEC-1998; 98US-0114308.
XX
PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
PI Kidd VJ, Lahti JM, Teitz T;
XX
DR WPI; 2000-452423/39.
XX
PT
```





sequence is a dinucleotide (TnCa) repeat. At positions 13639 and 13640 a single subclone disagrees with the sequence generated from the spanning PCR product. The subclone calls a T at both positions in contrast to the C called by the PCR product."

BASE COUNT 49306 a 40338 c 39825 g 48802 t  
ORIGIN

Query Match 82.9%; Score 17.4; DB 10; Length 178271;  
Best Local Similarity 94.7%; Pred. No. 3e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAGGGGATTCGAGATTGC 19  
|||||  
Db 80859 TAGGGGATTCGAGATTGC 80841

RESULT 14  
XLU07979

LOCUS 2742 bp DNA linear VRT 30-NOV-1995  
DEFINITION Xenopus laevis protein kinase (cdk2) gene, promoter and partial cds.

ACCESSION U07979.1 GI:473584

VERSION U07979.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

promoter

exon

5' UTR

CDS

intron

BASE COUNT

ORIGIN

788 a 559 c 667 g 728 t

1.2742

/organism="Xenopus laevis"

/db\_xref="taxon:8355"

/clone\_lib="genomic library of I. David (NIH, USA)"

/dev\_stage="adult"

1..2441

/gene="cdk2"

1..1928

/gene="cdk2"

/note="Using a cat reporter gene the presence of a promoter activity in this region was shown"

/evidence=experimental

1929..2441

/gene="cdk2"

1929..2325

/gene="cdk2"

2326..>2441

/evidence=experimental

2326..>2441

/gene="cdk2"

/codon\_start=1

/product="protein kinase"

/protein\_id="AA82133.1"

/db\_xref="GI:473585"

/translation="MENPKVEKIGEGTGVVYKARNRENGELVALKKIRLDT"

2442..>2742

/note="5' region of intron 1"

788 a 559 c 667 g 728 t

Query Match 80.0%; Score 16.8; DB 5; Length 2742;  
Best Local Similarity 90.0%; Pred. No. 4.7e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGGGATTCGAGATTGCGA 21  
|||||  
Db 2483 AGGGGATTCGAGATTGCGA 2502

RESULT 15  
AX281276 5518 bp DNA linear PAT 02-NOV-2001  
LOCUS AX281276  
DEFINITION Sequence 18 from Patent WO0177164.

ACCESSION AX281276

VERSION AX281276.1 GI:16608531

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

1383 a 142 c 1452 g 2541 t

1..5518

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="chemically treated genomic DNA (Homo sapiens)"

1383 a 142 c 1452 g 2541 t

Query Match 80.0%; Score 16.8; DB 6; Length 5518;

Best Local Similarity 90.0%; Pred. No. 5e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTCGAGATTGCG 20  
|||||

Db 46 TAGGGGATTCGAGATTGCG 65

Search completed: July 8, 2003, 04:02:36

Job time : 78.7495 secs

```

misc-feature      5169..5675
                  /note="similar to Homo sapiens EST AU144381
                  (NID:g1005902)"
misc-feature      5194..5761
                  /note="similar to Homo sapiens EST BF529656
                  (NID:g11617019)"
misc-feature      5237..6019
                  /note="similar to Homo sapiens EST AU130508
                  (NID:g10990862)"
misc-feature      5238..6043
                  /note="similar to Mus musculus EST BF166846
                  (NID:g11047198)"
misc-feature      5326..5853
                  /note="similar to Homo sapiens EST BG390376
                  (NID:g13283824)"
misc-feature      5361..6051
                  /note="similar to Homo sapiens EST AU135897
                  (NID:g10996436)"
misc-feature      5361..5812
                  /note="similar to EST BG862638 (NID:g14213176)"
misc-feature      5420..6048
                  /note="similar to Homo sapiens EST BE788113
                  (NID:g10209311)"
misc-feature      5551..6327
                  /note="similar to Homo sapiens EST AU130771
                  (NID:g10991125)"
misc-feature      5615..5984
                  /note="similar to Homo sapiens EST AA280659 (NID:g1923454)

Query Match      84.8% Score 17.8; DB 9; Length 181150;
Best Local Similarity 90.5%; Pred. NO.1.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY               1 TAGGGGATTGCGAGATTGCGA 21
                  ||||| ||||| |||||
Db               159132 TAGGGGACTCGAGACTGCGA 159112

RESULT 12
AX345696        AX345696      8951 bp      DNA      linear      PAT 01-FEB-2002
DEFINITION      Sequence 767 from Patent WO0200928.
ACCESSION       AX345696
VERSION         AX345696.1 GI:18493582
KEYWORDS        .
SOURCE          synthetic construct.
ORGANISM        synthetic construct
                artificial sequences.
REFERENCE       1
AUTHORS         Olek, A., Piepenbrock, C. and Berlin, K.
TITLE           Diagnosis of diseases associated with the immune system
JOURNAL         Patent: WO 0200928-A 767 03-JAN-2002;
                Epigenomics AG (DE)
FEATURES        Location/Qualifiers
                source          1..8951
                                /organism="synthetic construct"
                                /db_xref="taxon:32630"
                                /note="chemically created genomic DNA (Homo sapiens)"
BASE COUNT      1965 a      348 c      2538 g      3900 t
ORIGIN
Query Match      82.9% Score 17.4; DB 6; Length 8951;
Best Local Similarity 94.7%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY               3 GGGGATTGCGAGATTGCGA 21
                  ||||| ||||| |||||
Db               5581 GGGGTTTCGAGATTGCGA 5599

RESULT 13
AC091785/c      AC091785      178271 bp      DNA      linear      ROD 07-MAR-2002
LOCUS

```

```

DEFINITION      Genomic sequence for Mus musculus, clone RP23-349P1, complete
                  sequence.
ACCESSION       AC091785
VERSION         AC091785.5 GI:19224666
KEYWORDS        HTG.
SOURCE          Mus musculus.
ORGANISM        Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE       1 (bases 1 to 178271)
AUTHORS         McCormbie, W.R., Spiegel, L., de la Bastide, M., Preston, R.,
                Ferraro, K., Kuit, K., Nascimento, L., Zutavern, T., Ballifa, V.,
                Bell, M., Baker, J., Santos, L., Miller, B., Katzenberger, F.,
                Muller, S., King, L., Yang, C., Dike, S., Palmer, L., O'Shaughnessy, A.
                and Dedhia, N.
TITLE           Genomic sequence for Mus musculus, clone RP23-349P1, complete
                sequence
JOURNAL         Unpublished
REFERENCE       2 (bases 1 to 178271)
AUTHORS         McCormbie, W.R.
TITLE           Direct Submission
JOURNAL         Submitted (05-JUN-2001) Lita Annenberg Hazen Genome Sequencing
                Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
                Harbor, NY 11724, USA
REFERENCE       3 (bases 1 to 178271)
AUTHORS         McCormbie, W.R.
TITLE           Direct Submission
JOURNAL         Submitted (07-MAR-2002) Lita Annenberg Hazen Genome Sequencing
                Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
                Harbor, NY 11724, USA
COMMENT         On Mar 7, 2002 this sequence version replaced gi:16603978.
                This sequence was finished as follows unless otherwise noted: all
                regions were either double-stranded or sequenced with an alternate
                chemistry or covered by high quality data (i.e., phred quality >=
                30); an attempt was made to resolve all sequencing problems, such
                as compressions and repeats; all regions were covered by at least
                one plasmid subclone or more than one M13 subclone; and the
                assembly was confirmed by restriction digest.
                This submission represents a partial sequence of clone RP23-349P1.
                Only the unique (non-overlapping) sequence is represented. The
                remaining sequence overlaps clone RP23-469C2 (AC098839). An
                overlap of 2kb is included in this submission for reference. Base
                1 to base 2045 on RP23-349P1 overlaps the sequence of RP23-469C2
                from base 190433 to base 192478.

FEATURES        Location/Qualifiers
                source          1..178271
                                /organism="Mus musculus"
                                /db_xref="taxon:10090"
                                /clone="RP23-349P1"
                                /clone_lib="RPCT-23"
                                28760..28970
                                /note="We believe the assembly to be correct. The
                                sequence is covered by a single high quality subclone
                                sequenced with dye terminator chemistry. There is partial
                                coverage by three additional subclones with quality values
                                below phred30."
                                44235..44270
                                /note="We believe the assembly to be correct. The
                                sequence is covered by a single high quality subclone
                                sequenced with dye terminator chemistry. There is partial
                                coverage by an additional subclone with quality below
                                phred30."
                                125470..125510
                                /note="We believe the assembly to be correct. The
                                sequence is covered by a single high quality subclone
                                sequenced with dye terminator chemistry. There is partial
                                coverage by an additional subclone with quality below
                                phred30."
                                136300..136391
                                /note="We believe the assembly to be correct. The
                                sequence is covered solely by a PCR product which was
                                generated with a high fidelity polymerase. The sequence
                                is of high quality."
                                136305..136428
                                /note="We believe the assembly to be correct. The

```



```

hypothetical protein"
/codon_start=1
//transl_table=11
//product="conserved hypothetical protein"
//protein_id="AAK64751.1"
/db_xref="GI:14523156"
//translation="MEGSDGEQENLSRLDAQAOQTADPKTMOAKTKGPEERLRFSLTK
KDDWMTVGLAALALTFEGIOVILNLDMLDPAVLRLPMNVAKGGELIPIMLTVANAL
LEVLGILPVPSPRRNKLPRILVYVVALPAVLSLFLPMNVAAPVSGILSLILGAGG
ALQNPISFLIMVILIILAAKRYGRDRICDAGDVIDSYDPTLMREGGEVYLTMDTQ
PSGRILKPNSTVETIVNINSNWLPFPYVNEIKRPLASLELEVATNMEVYEQQ
GDIMQKVKIKRLISTPDELEGEKHPVHPFRSPENFWLTSLELVATPEAGRTQT
TRLIKEMLRNNAPEDRPVLPPKSLK"
complement(9248..10165)
//gene="Sma0179"
complement(9248..10165)
//gene="Sma0179"
complement(9248..10165)
//gene="Sma0179"
//function="Miscellaneous; Not classified regulator"
//note="Glimmer prediction; similar to glycine cleavage
system transcriptional activator, with bacterial LysR
regulatory helix-turn-helix protein domain"
/codon_start=1
//transl_table=11
//product="putative transcriptional activator"
//protein_id="AAK64752.1"
/db_xref="GI:14523157"

```

Query Match	84.8%;	Score 17.8;	DB 1;	Length 10253;
Best Local Similarity	90.5%;	Pred. No. 1.5e+02;		
Matches	19;	Conservative	0;	Mismatches 2;
			Indels	0;
			Gaps	0

```

QY      1 TAGGGGATTCGGAGATTGCGA 21
          | | | | | | | | | |
Db      368 TCGGGGATCCGGAGATTGCGA 388

```

RESULT	10
AC074016	
LOCUS	AC074016 179216 bp DNA linear HTG 09-AUG-2001
DEFINITION	Homo sapiens chromosome 2 clone RP11-155P18, WORKING DRAFT SEQUENCE, 2 unordered pieces.
VERSION	AC074016 AC074016
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 179216)
TITLE	Waterston,R.H.
JOURNAL	The sequence of Homo sapiens clone Unpublished
REFERENCE	2 (bases 1 to 179216)
AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUN-2000) Genome Sequencing Center, Washington University, School Of Medicine, 4444 Forest Park Parkway, St. Louis MO 63108, USA
COMMENT	On Aug 9, 2001 this sequence version replaced gi:19958173.

```

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H.N0155P18
-----
Summary Statistics
-----
Sequencing vector: M13; 48%
Sequencing vector: plasmid; 44%
Chemistry: Dye-Primer ET; 48% of reads
Chemistry: Dye-terminator Big Dye; 44% of reads
Assembly program: Phrap; version 0.950319

```

```

FEATURES
  source
    Consensus quality: 178616 bases at least Q40
    Consensus quality: 178729 bases at least Q30
    Consensus quality: 178811 bases at least Q20
    Insert size: 174000; agarose-fp
    Insert size: 181802; sum-of-contigs
    Quality coverage: 15.53 in Q20 bases; agarose-fp
    Quality coverage: 14.71 in Q20 bases; sum-of-contigs
    -----
    * NOTE: This is a 'working draft' sequence. It currently
    * consists of 2 contigs. The true order of the pieces
    * is not known and their order in this sequence record is
    * arbitrary. Gaps between the contigs are represented as
    * runs of N, but the exact sizes of the gaps are unknown.
    * This record will be updated with the finished sequence
    * as soon as it is available and the accession number will
    * be preserved.
    *
    * 1 82838: contig of 82838 bp in length
    * 82839 179216: gap of unknown length
    * 82839 179216: contig of 96278 bp in length.
    *
    Location/Qualifiers
      1..179216

```

	a	c	g	t	others
BASE COUNT	47524	37637	38145	55760	150
ORIGIN					

	Query Match	84.8%	Score 17.8	DB 2	Length 179216
	Best Local Similarity	90.5%	Ped. No. 1.8e+02		
	Matches 15, Conservative	0, Mismatches	2, Indels	0, Gaps	0
QY	1 TAGGGGATTGGAGATTGGCA	21			
Db	153057 TAGGGGACTGGAGACTGGCA	153077			

RESULT	11
FLOCUS	AC007256/c
DEFINITION	AC007256 Homo sapiens BAC clone RP11-575C6 from 2, complete sequence.
ACCESION	AC007256
VERSION	AC007256.5 GI:14327821
KEYWORDS	Htg.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens

REFERENCE	1 (bases 1 to 181150)
AUTHORS	Sulston,J.E. and Waterston,R.
TITLE	Toward a complete human genome sequence
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE	99063792
PUBMED	9847074
REFERENCE	2 (bases 1 to 181150)
AUTHORS	Du,F., Lapiant,Y., Doeberer,A. and Moore,B.
TITLE	The sequence of Homo sapiens BAC clone RP11-575C6
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 181150)
AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (07-Apr-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis MO 63108, USA
REFERENCE	4 (bases 1 to 181150)

JOURNAL Submitted (29-MAR-2001) Biological Sciences, Stanford University,  
371 Serra Mall, Stanford, CA 94305, USA  
Location/Qualifiers

FEATURES  
SOURCE

1..10253 "Sinorhizobium meliloti"

/organism="Sinorhizobium meliloti"

/strain="1021"

/db\_xref="taxon:382"

/plasmid="psyma"

/complement(123..308)

/gene="Sma0162"

/function="Miscellaneous; Unknown"

/note="glimmer prediction"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAK64743.1"

/db\_xref="GI:14523148"

/translation="MIDREKFEISTESPAPDAGILPMPVHRATTKOKNAKEPLDL

VWVPELHDTRRHVALS"

343..1572

/gene="piliQ2"

/note="Sma0163"

343..1572

/gene="piliQ2"

/function="Structural elements; Cell exterior; surface

structures"

/note="similar to hypothetical 44.3 kDa protein Y4XJ in

Rhizobium sp. NGR234"

/transl\_table=11

/codon\_start=1

/product="probable piliQ pilus assembly protein"

/protein\_id="AAK64744.1"

/db\_xref="GI:14523149"

/translation="MOREDELIVGDPRIATVPLTORSEFYILGNEIGSTSVTPDAEK

NPVGIIDIEVTDIKLSTIROSPSSVYKANSRATYISGNSADVAVATAEQA

SRKAGDEVAISIKITSSQOVLNRYVEINRDYKGLQISAIYKMSGSEYFNS

PRATSNTPAGSLISLIGESVDAIDALEDRMARLAEPNLARSGETSSFLAG

EFPPISEODGTIVSYKKEGVLDFTPTVLSDDLALDIEPEVSAIDNTASYRGI

AIPGFSEARATSVLDSGQSFMIAGLQSENNLTQRPGLGOLPILGALFSGKAQ

RREDDIVITVPHVKRPIDPLKVASPSDRTKRPTAEFFLGNDEVEVNGRGSASR

QARVRAPSSGHLELQ"

1569..1865

/gene="Sma0164"

1569..1865

/gene="Sma0164"

/function="Miscellaneous; Unknown"

/note="glimmer prediction"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAK64745.1"

/db\_xref="GI:14523150"

/translation="MIRSSRRPGFPGGLFASSILSGCQNHLEVRSETIALSGDDI

AANSVMQVDPMPRVRKOTSLATPADLEQYRPOQPMNEQNGNGETYPNPTTQ"

1887..2417

/gene="Sma0166"

1887..2417

/gene="Sma0166"

/function="Miscellaneous; Unknown"

/note="glimmer prediction"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAK64746.1"

/db\_xref="GI:14523151"

/translation="MNTTGASGRSVLRSQSGRRPELVMAIVATVILSCQSTSEVL

SGAEDPPSALASGSDGKSDLOGKLOFMNGNYGLAEKFRKAVELRODVAELMGL

AACVDRGRPLADRAVYQMLKVAQRPRITVNNNGSYQLAGEKAKAKAKLLERAA

PSGETIERNLALDLSR"

2501..3313

/gene="Sma0168"

CDS

2501..3313  
/gene="Sma0168"

/function="Macromolecule metabolism; Macromolecule

synthesis, modification; RNA synthesis, modification, DNA

transcription"

/note="glimmer prediction; with ribosomal RNA adenine

dimethylases domain"

/codon\_start=1

/transl\_table=11

/product="methyltransferase-like protein"

/protein\_id="AAK64747.1"

/db\_xref="GI:14523152"

/translation="MAMASDVLAASFGEARGLDQNYHTARAPPELVMLARRAC

LRKGISLIEIGAGTGLATERLEDRPHLLAVEPDRRLARLGRGDLKEDEVEITP

EKLIVPEKSELDVSAIAFAHWDAPALRRILRLRAGTAALEWVFGDVPDPFH

RATMFLSGHRTSPSGGTTKTPYGLNAGARLGEAEAGTADPELIDMTLADPPA

VRIVATYSNVTALPADREERLISGLEIAETEFAGVVTNNMTTSVYGRRE"

3374..4159

/gene="Sma0169"

3374..4159

/gene="Sma0169"

/function="Miscellaneous; Unknown"

/note="glimmer prediction"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAK64748.1"

/db\_xref="GI:14523153"

/translation="MALDRADYDAELARHNRQLVAADFGADRDVLDIGGAGGQTR

EAARAPQGEAIGVDISAMLEBARRSAAEGLRNAMPEQAGACFGPTGSEDCIS

RCQVMPFAPAPAAFNIGRAMRGARLVWQSRNENMSRAIRQALAPAVSAGA

ANFSLGDDPPVATDLISAGFTSIDPADQVEYFPGSYDAAFDLTSILYQDALAS

THEPPDRPLQRLDLEGHMTPGCVFDSRAWITARRAGGGG"

complement(4340..5104)

/gene="Sma0171"

complement(4340..5104)

/gene="Sma0171"

/function="Miscellaneous; Unknown"

/note="glimmer prediction"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAK64749.1"

/db\_xref="GI:14523154"

/translation="MNAKTLITLPLLAFAVYSIVSLYLSGRALLYPGASATPAE

RASWQNASIQTPDGETLHGLYSRGPSPVLFPGNDRVSNVGFQALAAAGIG

LIALSYRGYPGSSGTPNEHGLIDGIAADMLAARSGEITVVGOSLSEGVADTRAG

RPVAVYILVSAVLSVLAQTYVPPFVALTLKDPERSDLKLAGVRQPAVYVPRARH

HBIVFGRSSVSDRSRAQDAHLRCRPPRSVCGPGG"

5321..6241

/gene="Sma0172"

5321..6241

/gene="Sma0172"

/function="Miscellaneous; Hypothetical/Partial homology"

/note="glimmer prediction; C-terminal homology with

hypothetical protein"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAK64750.1"

/db\_xref="GI:14523155"

/translation="MSNSKDEVERIDMLEAELADITDEYELSEPTLSEKIRELYR

KAHPPALPRMDYFRALALQAEILKIDQWVYKQVYVFEGRDAKGGVIRKRTQ

RLNRPRIYVALPAPSDREKTOVYFORVYVLPAGEIVLEDSYNRKGVENRGAFA

TEBEVGEFPDVPDEPEMIIYRSGVIRYKWFSTDEEQOLRLIRIHDPLKQKLSPM

DLOGSRVMEVYTKAKESTFAKTRIREPNAIVANDKKRRLNCIDHLKQIYEDVP

HEIDITLPERIFNPEYKRVLPEDLYVPAKI"

complement(6637..7719)

/gene="Sma0175"

complement(6637..7719)

/gene="Sma0175"

/function="Miscellaneous; Hypothetical/Partial homology"

/note="glimmer prediction; partial homology with

TITLE Direct Submission  
JOURNAL Submitted (24-MAY-1996) D. Wallach, The Weizmann Institute, Dept of Membrane Research & Biophysics, Rehovot 76100, ISRAEL  
REMARK Revised by [3]  
REFERENCE 3 (bases 1 to 2887)  
AUTHORS Wallach, D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUN-1996) D. Wallach, The Weizmann Institute, Dept of Membrane Research & Biophysics, Rehovot 76100, ISRAEL  
REFERENCE 4 (bases 1 to 2887)  
AUTHORS Mandruzzato, S., Brasseur, F., Andry, G., Boon, T. and van der Bruggen, P.  
TITLE A CASP-8 mutation recognized by cytolytic T lymphocytes on a human head and neck carcinoma  
J Exp Med 186 (5), 785-793 (1997)  
JOURNAL 97419196  
MEDLINE 9271594  
PUBMED  
FEATURES  
Source  
Location/Qualifiers  
1. 2887  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="thymus"  
292. 1731  
/function="protease"  
/note="triggers cell death"  
/product="MACH-alpha-1"  
/protein\_id="CAA6653.1"  
/db\_xref="GI:2440071"  
/db\_xref="SWISS-PROT:Q14790"  
/translation="MDFSRNLYDIGEOLDSLSLKLFLSDYIPQKOEPIKDALML  
FORLQEKMLERSEN,SLFLKELFLRINRDLITYLNTPEKMERELQTPGAQISAYR  
VLYXOISEVSRSELRSFLQELRISKLDLDDNLDIFLEMKRVYLIGEKLDIL  
KVCQAIKNSLIKIIINDYEESEKRSSESLSGDPDFSGEELCGVMTSDSPRODSE  
SOTLKYVQMKKPRGVCYLIINHNFAKAREVPELHTRNGKHITDAGATITTFEE  
LHFEIKPHDCTVEQIYELIKTYLQMLDSSNMDCITCCILSHDGKIIITGTGQEARPI  
ELTFSQFTGLKCPSLAGKPRVFFIQACQGDNYOKGIPVETDSEQYLETMDSSPTRY  
IPDEADFLGNATVNNCVSRNPAGVTWYIOSLCOSLRERCPRGDIITLITLEVNEY  
SNKDRKNNKQKMPQPTFLRKILYPSD"

BASE COUNT 853 a 595 c 625 g 814 t  
ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 2887;  
Best Local Similarity 90.5%; Pred. No. 1.4e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTCGGAGATTGCGA 21  
221 TAGGGGACTCGGAGACTGCGA 241

RESULT 8  
AB038980S2 6502 bp DNA linear PRI 23-FEB-2001  
LOCUS Homo sapiens gene for caspase-8, exon 3, exon 4.  
DEFINITION  
ACCESSION AB038981.1 GI:12862688  
VERSION  
KEYWORDS 2 of 6  
SEGMENT Homo sapiens DNA.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
AUTHORS Hadano, S., Yanagisawa, Y., Skaug, J., Fischer, K., Nasir, J.,  
Martindale, D., Koop, B.F., Scherer, S.W., Nicholson, D.W.,  
Roulean, G.A., Ikeda, J.-E. and Hayden, M.R.  
TITLE Cloning and characterization of three novel genes, ALS2CR1,  
ALS2CR2, and ALS2CR3, in the juvenile amyotrophic lateral sclerosis  
(ALS2) critical region at chromosome 2q33-q34: candidate genes for  
ALS2  
JOURNAL Genomics 71 (2), 200-213 (2001)  
MEDLINE 21100893

REFERENCE 2 (bases 1 to 6502)  
AUTHORS Hadano, S., Ikeda, J. and Hayden, M.R.  
TITLE Direct Submission  
JOURNAL Submitted (24-FEB-2000) Shinji Hadano, Tokai University, The  
Institute of Medical Sciences, Bohseidai, Isehara, Kanagawa  
259-1193, Japan (E-mail: shinji@nga.med.u-tokai.ac.jp,  
Tel:+81-463-91-5095, Fax:+81-463-91-4993)  
FEATURES  
Source  
Location/Qualifiers  
1. 6502  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2q33"  
/map="2q33"  
1. 1478  
/note="intron is approximately 24.0 kbp long"  
/number=2  
1479. 1743  
/product="caspase-8"  
/note="CDS is reported in Acc# AB038985  
alternative 5'UTR"  
/number=3  
1744. 3924  
/number=3  
3925. 3970  
/product="caspase-8"  
/note="CDS is reported in Acc# AB038985  
alternative 5'UTR"  
/number=4  
3971. >6502  
/note="intron is approximately 5.9 kbp long"  
/number=4

BASE COUNT 1718 a 1361 c 1504 g 1919 t  
ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 6502;  
Best Local Similarity 90.5%; Pred. No. 1.4e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTCGGAGATTGCGA 21  
1699 TAGGGGACTCGGAGACTGCGA 1719

RESULT 9  
AE007203 10253 bp DNA linear BCT 15-AUG-2001  
LOCUS Sinorhizobium meliloti plasmid pSyma section 9 of 121 of the  
DEFINITION complete plasmid sequence.  
ACCESSION AE007203 AE006469  
VERSION AE007203.1 GI:14523147  
KEYWORDS  
SOURCE  
ORGANISM Sinorhizobium meliloti.  
Sinorhizobium meliloti.  
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
Rhizobiaceae; Sinorhizobium.  
1 (bases 1 to 10253)  
Barnett, M.J., Fisher, R.F., Jones, T., Komp, C., Abola, A.P.,  
Barloy-Hubler, F., Bowser, L., Capela, D., Galibert, F., Gouzy, J.,  
Gurjal, M., Hong, A., Huizart, L., Hyman, R.W., Kahn, D., Kahn, M.L.,  
Kallman, S., Keating, D.H., Palm, C., Peck, M.C., Surzycki, R.,  
Wells, D.H., Yeh, K.-C., Davis, R.W., Federspiel, N.A. and Long, S.R.  
TITLE Nucleotide sequence and predicted functions of the entire  
Sinorhizobium meliloti pSyma megaplasmid  
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)  
JOURNAL 21396509  
MEDLINE 11481432  
PUBMED

REFERENCE  
AUTHORS Barnett, M.J., Fisher, R.F., Jones, T., Komp, C., Abola, A.P.,  
Barloy-Hubler, F., Bowser, L., Capela, D., Galibert, F., Gouzy, J.,  
Gurjal, M., Hong, A., Huizart, L., Hyman, R.W., Kahn, D., Kahn, M.L.,  
Kallman, S., Keating, D.H., Palm, C., Peck, M.C., Surzycki, R.,  
Wells, D.H., Yeh, K.-C., Davis, R.W., Federspiel, N.A. and Long, S.R.  
TITLE Direct Submission

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2503)  
 Breckenridge, D.G., Nguyen, M., Kuppig, S., Reth, M. and Shore, G.C.  
 The procaspase-8 isoform, procaspase-8L, recruited to the BAP1  
 complex at the endoplasmic reticulum  
 Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4331-4336 (2002)

JOURNAL 21927603  
 MEDLINE 11917123  
 PUBMED 11917123

REFERENCE 2 (bases 1 to 2503)  
 Breckenridge, D.G., Nguyen, M., Kuppig, S., Reth, M. and Shore, G.C.  
 Direct Submision  
 Submitted (24-SEP-2001) Biochemistry, McGill University, 3655  
 Promenade Sir William Osler, Montreal, PQ H3G1Y6, Canada

AUTHORS  
 JOURNAL

FEATURES  
 Location/Qualifiers  
 1..2503  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..2503  
 /gene="CASP8"  
 115..1731  
 /gene="CASP8"  
 /note="caspase precursor; alternatively spliced"  
 /codon\_start=1  
 /product="procaspase-8L"  
 /protein\_id="AAL87628.1"  
 /db\_xref="GI:19401519"  
 /translation="MEGRARAVIESRRNPFILGAPTPPAHVELGRGDEETAV  
 PKRGADYILPEKMPFSRNIDIGQDSEDLASLKFSLDIIPQRKEPKDALM  
 LFORQERMLEESNLSFLKELRIRKLDLITYLNTREEMERELQTPGKQISAY  
 RVMYQISEVSRSELSFKELRIRKLDLITYLNTREEMERELQTPGKQISAY  
 LKRVQAINSLKLIINDYEFSEKSSSESGPDESNCEELGQWTLSDSPRODS  
 ESQILDVYQKSKPRGICLIINHNNAKAREVKPKHSIRDRNGTILDGATTFEE  
 ELHEIRPHDCTVEQIYEILKITYQLMDHSMNDFICCIISHGKGIITGDSQEAR  
 YELNSOPTGLKCPSLAGKPKVFFIQACGQNYGKIPVEDSEQPLEMDISSPOR  
 YIPEDADFLGMAVNNVCYSRNPAEGTWIQSLCSLRERCPRGDDILITLEVNT  
 VSNDDKRNKGKMPQPTFLRKLVPPSD"

BASE COUNT 726 a 528 c 593 g 656 t

ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 2503;  
 Best Local Similarity 90.5%; Pred. No. 1.3e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TAGGGATTGGAGATTGCGA 21  
 ||||| ||||| ||||| |||||  
 Db 221 TAGGGACTCGAGACTGCGA 241

RESULT 5  
 AR211527 2887 bp DNA linear PAT 20-JUN-2002

LOCUS AR211527  
 DEFINITION Sequence 14 from patent US 6399337.  
 ACCESSION AR211527  
 VERSION AR211527.1 GI:21514872  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 Unclassified.

REFERENCE 1 (bases 1 to 2887)  
 Wallach, D., Boldin, M., Goncharov, T. and Golsteyn, V.V.  
 Modulators of the function of Fas receptors and other proteins  
 Patent: US 6399337 A 14 04-JUN-2002;  
 Location/Qualifiers  
 1..2887  
 /organism="unknown"

BASE COUNT 853 a 595 c 625 g 814 t

ORIGIN

Query Match 84.8%; Score 17.8; DB 6; Length 2887;  
 Best Local Similarity 90.5%; Pred. No. 1.4e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TAGGGATTGGAGATTGCGA 21  
 ||||| ||||| ||||| |||||  
 Db 221 TAGGGACTCGAGACTGCGA 241

RESULT 6  
 AX134419 2887 bp DNA linear PAT 29-MAY-2001

LOCUS AX134419  
 DEFINITION Sequence 15 from Patent WO0129232.  
 ACCESSION AX134419  
 VERSION AX134419.1 GI:14271065  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2887)  
 Cordell, B. and Li, Y.  
 Functional cloning of genes encoding proteins/enzymes involved in  
 proteolytic cleavage  
 Patent: WO 0129232-A 15 26-APR-2001;  
 Scios Inc. (US)  
 Location/Qualifiers  
 1..2887  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 292..1731  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAC39526.1"  
 /db\_xref="GI:14271066"

BASE COUNT 853 a 595 c 625 g 814 t

ORIGIN

Query Match 84.8%; Score 17.8; DB 6; Length 2887;  
 Best Local Similarity 90.5%; Pred. No. 1.4e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TAGGGATTGGAGATTGCGA 21  
 ||||| ||||| ||||| |||||  
 Db 221 TAGGGACTCGAGACTGCGA 241

RESULT 7  
 HSMACHAL 2887 bp mRNA linear PRI 25-SEP-1997

LOCUS HSMACHAL  
 DEFINITION H.sapiens mRNA for MACH-alpha-1 protein.  
 ACCESSION X98172  
 VERSION X98172.1 GI:1403318  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2887)  
 Boldin, M.P., Goncharov, T.M., Golsteyn, V.V. and Wallach, D.  
 Involvement of MACH, a novel MORT1/FADD-interacting protease, in  
 Fas/APO-1- and TNF receptor-induced cell death  
 Cell 85 (6), 803-815 (1996)

JOURNAL 96279826  
 MEDLINE 8681376  
 PUBMED 8681376

REFERENCE 2 (bases 1 to 2887)  
 Wallach, D.

complex at the endoplasmic reticulum  
Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4331-4336 (2002)

JOURNAL 21927603  
MEDLINE 11917123  
PUBMED 2 (bases 1 to 777)  
REFERENCE Breckenridge,D.G., Nguyen,M., Kupfig,S., Reth,M. and Shore,G.C.  
AUTHORS Submitted (24-SEP-2001) Biochemistry, McGill University, 3655  
TITLE Promenade Sir William Osler, Montreal, PQ H3G1V6, Canada  
JOURNAL Location/Qualifiers

FEATURES  
source 1..777  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="4"  
1..>777  
/gene="CASP8"  
266..>777  
/gene="CASP8"  
/note="caspase precursor; alternatively spliced"  
/codon\_start=1  
/product="procaspase-8"  
/protein\_id="AA187632.1"  
/db\_xref="GI:19401530"  
/translation="MDFSRNLYDIGEQLDSEDLASLFLSLDYTPQRKQEPKDALML  
FORLOEKRMLESNLSPLKELLRINRDLITLNTKREMERLOTGPRAQISAYR  
VMLOYISEVSRSELSRFLQSEISCKLDDMLDIFIEMEKRVIIIGEGKLDIL  
KRVCAQINKSL"  
KRVCAQINKSL"

BASE COUNT 218 a 156 c 209 g 194 t

ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 777;  
Best Local Similarity 90.5%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTCGAGATTGCGA 21  
|||||  
221 TAGGGACTCGAGACTGCGA 241

Db

RESULT 2  
AF210257 982 bp DNA linear PRI 06-JUN-2000  
LOCUS Homo sapiens caspase 8 gene, 5' flanking region.  
DEFINITION AF210257  
ACCESSION AF210257.1 GI:8272474  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 982)  
Teitz,T., Wei,T., Valentine,M.B., Vanin,E.F., Grenet,J.,  
Valentine,V.A., Behm,F.G., Look,A.T., Lantti,J.M. and Kidd,V.J.  
caspase 8 is deleted or silenced preferentially in childhood  
neuroblastomas with amplification of MYCN

REFERENCE  
AUTHORS 1 (bases 1 to 982)  
Teitz,T., Wei,T., Valentine,M.B., Vanin,E.F., Grenet,J.,  
Valentine,V.A., Behm,F.G., Look,A.T., Lantti,J.M. and Kidd,V.J.  
Direct Submission  
Submitted (01-DEC-1999) Tumor Cell Biology, St. Jude Children's  
Research Hospital, 332 N. Lauderdale, Memphis, TN 38105, USA  
JOURNAL Location/Qualifiers

FEATURES  
source 1..982  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2q33-q34"  
1..982  
/note="5' flanking region of caspase 8 gene"

BASE COUNT 198 a 235 c 241 g 308 t

misc.feature

ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 982;  
Best Local Similarity 90.5%; Pred. No. 1.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTCGAGATTGCGA 21  
|||||  
536 TAGGGACTCGAGACTGCGA 556

Db

RESULT 3  
AF380342 1723 bp mRNA linear PRI 03-JUN-2002  
LOCUS Homo sapiens caspase-8L mRNA, complete cds.  
DEFINITION AF380342  
ACCESSION AF380342.1 GI:14211397  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1723)  
Himeji,D., Horiuchi,T., Tsukamoto,H., Hayashi,K., Watanabe,T. and  
Harada,M.  
Characterization of caspase-8L: a novel isoform of caspase-8 that  
behaves as an inhibitor of the caspase cascade  
Blood 99 (11), 4070-4078 (2002)

JOURNAL 22005982  
MEDLINE 12010809  
PUBMED 2 (bases 1 to 1723)  
REFERENCE Himeji,D., Tsukamoto,H. and Horiuchi,T.  
AUTHORS Direct Submission  
TITLE Submitted (10-MAY-2001) Medicine and Biosystemic Science, Kyushu  
University Graduate School of Medical Science, 3-1-1 Maidashi,  
Higashi-ku, Fukuoka 812-8582, Japan  
JOURNAL Location/Qualifiers

FEATURES  
source 1..1723  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="peripheral blood lymphocyte"  
86..916  
/codon\_start=1  
/product="caspase-8L"  
/protein\_id="AAK57457.1"  
/db\_xref="GI:14211398"  
/translation="MDFSRNLYDIGEQLDSEDLASLFLSLDYTPQRKQEPKDALML  
FORLOEKRMLESNLSPLKELLRINRDLITLNTKREMERLOTGPRAQISAYR  
VMLOYISEVSRSELSRFLQSEISCKLDDMLDIFIEMEKRVIIIGEGKLDIL  
KRVCAQINKSLKINLYEFSEKRSSESGSDPDSNGELGQWVITSDSPQDSE  
SQTLDKVYQMKSRPRGICLTINHNFRKAREKVPKLSIRNGTHLDAGIVEPKREK

BASE COUNT 538 a 344 c 411 g 430 t

ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 1723;  
Best Local Similarity 90.5%; Pred. No. 1.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTCGAGATTGCGA 21  
|||||  
15 TAGGGACTCGAGACTGCGA 35

Db

RESULT 4  
AF422925 2503 bp mRNA linear PRI 03-APR-2002  
LOCUS Homo sapiens procaspase-8L (CASP8) mRNA, complete cds.  
DEFINITION AF422925  
ACCESSION AF422925.1 GI:19401518  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens.



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:07:09 ; Search time 74.5828 Seconds

(Without alignments)  
8194.363 Million cell updates/sec

Title: US-09-477-082-29

Perfect score: 21

Sequence: 1 tagggagtcgagcattgcga 21

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: gb\_da:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pin:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_hgo\_hum:\*  
40: em\_hgo\_mus:\*  
41: em\_hgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17.8	84.8	777	9	AF422929	AF422929 Homo sapi
2	17.8	84.8	982	9	AF210257	AF210257 Homo sapi
3	17.8	84.8	1723	9	AF380342	AF380342 Homo sapi
4	17.8	84.8	2503	9	AF422925	AF422925 Homo sapi
5	17.8	84.8	2887	6	AR211577	AR211577 Sequence
6	17.8	84.8	2887	6	AX134419	AX134419 Sequence
7	17.8	84.8	2887	6	HSMACHA1	X98172 H. sapiens m
8	17.8	84.8	6502	9	AB038980S2	AB038981 Homo sapi
9	17.8	84.8	10253	1	AE007203	AE007203 Simorhizo
10	17.8	84.8	179216	2	AC074016	AC074016 Homo sapi
11	17.8	84.8	181150	5	AC007256	AC007256 Homo sapi
12	17.4	82.9	8651	6	AX345696	AX345696 Sequence
13	17.4	82.9	178271	10	AC091785	AC091785 Genomic S
14	16.8	80.0	2742	5	XL007979	U07879. Xenopus lae
15	16.8	80.0	5518	6	AX281276	AX281276 Sequence
16	16.8	80.0	5518	6	AX344611	AX344611 Sequence
17	16.8	80.0	5518	6	AX345119	AX345119 Sequence
18	16.8	80.0	141382	2	AC097575	AC097575 Rattus no
19	16.8	80.0	148783	2	CNS0808K	AL771879 Oryza sat
20	16.8	80.0	151944	2	AC115475	AC115475 Rattus no
21	16.4	78.1	2785	3	AB007406	AB007406 Ha10cynth
22	16.4	78.1	38687	3	U23515	U23515 Caenorhabd1
23	16.4	78.1	70132	9	AC091938	AC091938 Homo sapi
24	16.4	78.1	116211	8	AP003843	AP003843 Oryza sat
25	16.4	78.1	144561	2	AP004307	AP004307 Oryza sat
26	16.4	78.1	150699	2	AC099245	AC099245 Rattus no
27	16.4	78.1	179686	2	AC023987	AC023987 Homo sapi
28	16.4	78.1	191052	10	AL683814	AL683814 Mouse DNA
29	16.2	77.1	290	3	HS21H7F	Z60409 H. sapiens C
30	16.2	77.1	1759	3	AF177464	AF177464 Drosophila
31	16.2	77.1	7618	1	AE005734	AE005734 Ca1obact
32	16.2	77.1	20303	3	AC006749	AC006749 Caenorhab
33	16.2	77.1	59527	2	AC014163	AC014163 Drosophila
34	16.2	77.1	63268	2	AC013683	AC013683 Homo sapi
35	16.2	77.1	126631	8	AP003379	AP003379 Oryza sat
36	16.2	77.1	133084	2	AC099208	AC099208 Rattus no
37	16.2	77.1	171705	2	AC104166	AC104166 Homo sapi
38	16.2	77.1	173536	2	AC074388	AC074388 Homo sapi
39	16.2	77.1	176145	2	AL731550	AL731550 Homo sapi
40	16.2	77.1	176444	2	AC130018	AC130018 Rattus no
41	16.2	77.1	184271	3	AC009256	AC009256 Drosophila
42	16.2	77.1	186184	2	AC107275	AC107275 Rattus no
43	16.2	77.1	191709	2	AC097846	AC097846 Rattus no
44	16.2	77.1	193444	8	AF229199	AF229199 Oryza sat
45	16.2	77.1	193603	2	AC093932	AC093932 Rattus no

## ALIGNMENTS

RESULT 1  
AF422929  
LOCUS Homo sapiens clone 4 procaspase-8 (CASP8) mRNA, partial cds;  
DEFINITION  
ACCESSION AF422929  
VERSION AF422929.1 GI:19401529  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 777)  
AUTHORS Breckenridge,D.G., Nguyen,M., Kuppig,S., Reth,M. and Shore,G.C  
TITLE The procaspase-8 isoform, procaspase-8L, recruited to the BAP31



```

?      NAME/KEY: CDS
?      LOCATION: (24168)...(24307)
?      NAME/KEY: CDS
?      LOCATION: (23696)...(25908)
?      NAME/KEY: CDS
?      LOCATION: (27235)...(27246)
?      US-09-800-631-96

Query Match      4.64; Score 34.8; DB 10; Length 30310;
Best Local Similarity 50.08; Pred. No. 5.4;
Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY      555 CTTTCCCTCTTTCGCGCTCTGGAAGGGGTTTCCTTTATGTCTTCACCCGCCCTTTC 614
        ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      13340 CTTTCTCTTGTGTGTGTCTCTCTCCCTCTTCATTTCTGTGTGTCTCTCCCCACCT 13399

QY      615 CCTCTCCTGCCCTCTGTTTGTGTGCCCCAAAAACAAGTTCTCTAAAGCTTTTCATG 674
        ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      13400 TCCGCTCCACCTCCCTCTTATCTGTGCCCAACTGCACCTCCCTTATTTACAAATTAA 13455

QY      675 GATTCCGGGGAATAATTACCTGCACCCGTTTGCAAAATGAAGCTTTTGTTTTGA 728
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      13460 GGTGGGAAGCAGGTTTACAGGAATACCTCGAGTGTCTGACTATGTTCTTGTCTCA 13513

```

```

1      RESULT 13
2      US-10-105-637-4
3      :
4      : Sequence 4, Application US/10105637
5      :
6      : Publication No. US20030087252A1
7      :
8      : GENERAL INFORMATION:
9      :
10     : APPLICANT: Morris, David
11     : APPLICANT: Engelhard, Eric
12     : TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS IN CANCER
13     : TITLE OF INVENTION: ASSOCIATED WITH ALTERED EXPRESSION OF PDM1
14     : FILE REFERENCE: 529452500120
15     : CURRENT APPLICATION NUMBER: US/10/105,637
16     : CURRENT FILING DATE: 2002-07-23
17     : PRIOR APPLICATION NUMBER: US 10/034,650
18     : PRIOR FILING DATE: 2001-12-20
19     : PRIOR APPLICATION NUMBER: US 09/747,377
20     : PRIOR FILING DATE: 2000-12-22
21     : PRIOR APPLICATION NUMBER: US 09/798,586
22     : PRIOR FILING DATE: 2001-03-02
23     : NUMBER OF SEQ ID NOS: 7
24     : SOFTWARE: FASTSEQ for Windows Version 4.0
25     : SEQ ID NO 4
26     :
27     : LENGTH: 63720
28     :
29     : TYPE: DNA
30     :

```

```

Query Match          4.6%: Score 34.8; DB 9; Length 63720;
Best Local Similarity 52.8%: Pred. No. 8;
Matches 75; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY      266 GCTGCCTGAGCCAGGCTCTCTGTGTGTTCTCTCTGTAGCCGATGCCCTTGTACTTGTCTA 325
          ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      38465 GCTGCTGAGCCGCCCACTTTCTGTGTGCTGCTCTCAACCTCCCTCCATCCACTCTGAGCTG 38528
          ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY      326 CTTTTCACGTGTAGCAGCTCTCCAGTTCCTCTGTCTGTACTCTTTTGTGTCTCCATCAAGCTCCCT 385
          ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      38525 CTCCGCCACCGCGCTCCCCACCCCTCTCTCCCTGTGTACTTGTCCCTTCCCTTTCCTCTCT 38588
          ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY      386 GCCGCGCTCGAATGCAGATACAC 407
          ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      38585 ACTTCCGTGTGTCTCACAATGCAC 38606
          ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-10-123-155-210/c
; Sequence 210, Application US/10123155
; Publication No. US20030086794A1
; GENERAL INFORMATION:

```



Best Local Similarity 27.9%; Pred. No. 0.39;  
Matches 97; Conservative 77; Mismatches 172; Indels 2; Gaps 1;

```

QY 319 TTTCCTACTTTTACTCTGAGCAGTCTCCAGTCTCTGCTGACCTTTTGTCTCCAG 378
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1441 TTTTTC.RCTTTTTRCBACSDYDAAYDAYSTYDARCA..AKDASTYDARCBTYYSR 1382
QY 379 CTTCCTGCGCCGCGGATGAGATGACAGGACTCCCTCTGTGAGACCCCTTGAGAGT 438
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1381 CTTRCTTRC.RC.CCYSTTTNSTNSTYASRCBTTRCDAIB.TTRC..TTTTYTTT 1322
QY 439 CCAGAAGACTTATCAATCACTTTTCTTTTCTTTTCAATTTGGCCCTGGGGCCGAGGT 488
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1321 BCTBRCNCTCTCYTTRCYRA.TTTTYSYSTTCTTTTCTTTTCTTTTYSNCTB.AYDA. 1262
QY 499 TAAGACTTATCTGTCATCTCTGCAATGACAAAGCCCTGAGGCGACAGCCCTTT 558
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1261 .YSSRRCNCTBACTTATTRCRCSR.ACAAYSRCBTBNSTYRACAKATTR--CRCTT. 1204
QY 559 CCCCCTCTTCGCGCTGAGGAGGCTTCCCTTTATGCTTCCACCCCTTTCCCT 618
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1203 RCTBTTRTTR..CABTBTTRCTTYDA.TTTTADATTTARCTTRCCTT.TTCYT 1144
QY 619 CCGTGCCTCTGTTTGTGTTCCCAAAAACAGTTCTTAACGTTT 666
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1143 RCSRRCNCTCYTAKYSTY..SAKRAC.SYSRCTYTDADYSTDY 1096

```

## RESULT 6

```

US-10-198-846-7264/C
; Sequence 7264, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongzao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinhmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7264
; LENGTH: 901
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 280, 324, 379, 383, 395, 397, 399, 407, 413, 427, 431, 432,
; LOCATION: 433, 445, 458, 465, 476, 479, 499, 512, 519, 520, 526, 527,
; LOCATION: 529, 533, 540, 541, 559, 576, 577, 603, 604, 605, 609, 610,
; LOCATION: 617, 623, 624, 628, 649, 659, 660, 667, 672, 679
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 680, 682, 684, 685, 686, 687, 688, 689, 711, 712, 713, 714,
; LOCATION: 715, 718, 721, 723, 740, 751, 752, 755, 756, 757, 758,
; LOCATION: 759, 760, 761, 772, 775, 779, 783, 784, 787, 788, 789, 790,
; LOCATION: 792, 794, 795, 797, 802, 809, 814, 815, 816, 817, 818
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 819, 820, 821, 822, 828, 830, 831, 832, 833, 834, 836, 839,
; LOCATION: 840, 841, 842, 843, 845, 847, 849, 851, 855, 857, 862,
; LOCATION: 865, 868, 869, 886, 900, 901
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-7264

```

Query Match 4.8%; Score 36; DB 9; Length 901;  
Best Local Similarity 42.1%; Pred. No. 0.32;  
Matches 69; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

```

QY 584 TTCTCTTATGCTTCCACCCCAACCTTCCCTCCCTCCCTGCTGTTTGTGCCCCA 643
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 778 TTTTNTTTTTTTTTTTTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCC 719
QY 644 AAAAAGAGTCTGCTTAAGCTTTGATGTGATGTGCGGAATTAACCTGACCCGTT 703
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 718 NCCNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTNNNNNNNNNNNNNNNNNNNNNN 659
QY 704 TGCATAATGACCTTTTGTGATGCTGTAACCTGATGTTT 747
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 658 AAAAATAATTTTTTTTTTTTTTTTTTTTTTTTCCNCCNNGGATWTT 615

```

## RESULT 7

```

US-09-764-891-7205
; Sequence 7205, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7205
; LENGTH: 9743
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7205

```

Query Match 4.8%; Score 35.8; DB 9; Length 9743;  
Best Local Similarity 53.1%; Pred. No. 1.4;  
Matches 76; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```

QY 246 CTCTGAGGACCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 305
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1212 CCTGTGAGTGACACCTGTTGTGCTTGCCCTGCGACAGGAGAGATGCTCATCTT 1271
QY 306 CGATGCTTTGACTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 365
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1272 GGAGTTCAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1331
QY 366 TTGTCTCTCAAGCTTCCCTGCC 388
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1332 TATGAGCCGCGGCTTCTTCC 1354

```

## RESULT 8

```

US-10-295-403-75/C
; Sequence 75, Application US/10295403
; Publication No. US20030101481A1
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, Lynne
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Zhang, James
; APPLICANT: Benito, Maria-Ines
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Fromm, Mike
; TITLE OF INVENTION: PLANT GENE SEQUENCES I
; FILE REFERENCE: MBI-0003
; CURRENT APPLICATION NUMBER: US/10/295,403

```

Db	365	Y....Y..SBNC	TTB.W.DSCM..R....YHBA.TH.SSDGM.C.A..YVC...SRSHM	306
QY	437	GTCCAGACAGC	TTTATCAATCCACTTTTCTTTCTTTTATTTGAGCCGGGGCCGACG	496
Db	305	A..HA..MC...BTBSOC..MCR.S..BT..BHMS..THVC.YMR.R..TK..TY..T.SDA..C..		246
QY	497	GTTAAGACTT	TTATCTGNCATCTCTGCAATGCAAGATGCCCTGAGTGCACAGGCCCT	556
Db	245	ADC..DT.BB..NXYIYS..SWY.H.H.DT.CYX...TDT.TCC.CH.TNC..CHNDHC...		186
QY	557	TTCCCTCTTTGCGCGCTGCGAGAGGGGTTTCCTTTATGTCTTCCACCCACCCTTTCC		616
Db	185	A..Y..HT..CAT..T.W....TSBNSM....TYBTCTYBT....D..MNB.YRS.D.MRMNMC.		126
QY	617	CTCCCTGCCCTCTGTTTGTGTTGTCGCCAAAACAAAGTTCGTAACGTTTGAGATGGA		676
Db	125	MT..YMYTKNAYDT..HS.Y.TWTRSNNAHBCGY.YAK.Y.ATYH.BCSB...SBW...		66
QY	677	TTGCGGGAAATTAACCTGCACCCGTTTGCAAAATGAACTTTTTTTTGATCCTGTAC		736
Db	65	TY..TYS.YASA..TM.B..N.RCRM...T.T.MY...CCY.T..TH.DCW....TY...C		6
QY	737	ACTGG	741	
Db	5	BYTBK	1	
<p>RESULT 4  US-09-759-359A-3  : Sequence 3, Application US/05759359A  : Patent No. US2002094560A1  : GENERAL INFORMATION:  : APPLICANT: ABU-THREIDEH, Jane et al  : TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  : TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  : TITLE OF INVENTION: THEREOF  : FILE REFERENCE: CI001043  : CURRENT APPLICATION NUMBER: US/09/7759, 359A  : CURRENT FILING DATE: 2001-01-16  : NUMBER OF SEQ ID NOS: 3  : SOFTWARE: FastSeq for Windows Version 4.0  : SEQ ID NO 3  : LENGTH: 90541  : TYPE: DNA  : ORGANISM: Human  : US-09-759-359A-3</p>				
<p>Query Match 5.1%; Score 38.4; DB 10; Length 90541;  Best Local Similarity 44.5%; Pred. No. 0.63;  Matches 153; Conservative 0; Mismatches 191; Indels 0; Gaps 0;</p>				
QY	285	CTGTGTGTGTTCTCTCTGAGCCGATGCTTTGACTTTGCTACTTTTACCTCTGAGCAGT		344
Db	34667	CTGATTTCTTCCTCTCCCTTCCGTCGCTCCCTCCCTCCCTCCCTCCCTTCTTCTTTCTTT		34722
QY	345	CTCCAGTTCCTCTGCTACTCTTTTGTCTCCAGACTCTCCGCGCGCGCGATGACAGATA		404
Db	34727	TTTCTTTTCCCTGCGTCGCTTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTC		34788
QY	405	CACGGACTCCCTGTGTGACCCGTTTGGAGAGTCCAGAACTTATTCATCACCCTTT		464
Db	34787	TTTCTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTC		34844
QY	465	TTTCTTTTTCATTTGCGCCCGGGGGCGCGAGGTAGTACTTATCTGTCTGATCTGTC		524
Db	34847	TTTCTTTCTTCTCTTTTCTTTCTTTCTTTCTTCTCTTCTTCTTCTTCTTCTTCTTCTTC		34900
QY	525	GAATACAGAAATGCCCTGAGTGCACAGACCCCTTTCCTCTTGGCTGCTGAAGGGAT		584
Db	34907	TTTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTC		34966
QY	585	TCGCTTTATGTCTTCACACCCGCTTTCCTCCCTCCCTGCGCCTC		628

```

Db      34967 TCTCCTTTCTCCCTCTCTCTCCTCCCTCCCTCCCTCCCTCCCTC 35010
          || | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 5
US-10-242-056-60/c
Sequence 60, Application US/10242056
Publication No. US20030113323A1
GENERAL INFORMATION:
APPLICANT: Ensign, Jerald C
APPLICANT: Bowen, David J
APPLICANT: Petell, James
APPLICANT: Fatig, Raymond
APPLICANT: Schoonover, Sue
APPLICANT: french-Constant, Richard
APPLICANT: Orr, Gregory L
APPLICANT: Merlo, Donald J
APPLICANT: Roberts, Jean L
APPLICANT: Rocheteau, Thomas A
TITLE OF INVENTION: Insecticidal Protein Toxins from
TITLE OF INVENTION: Photornabads
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESSER: DOWELANCO
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: IN
COUNTRY: US
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/242,056
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/063,615
FILING DATE: 18-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/395,497
FILING DATE: 28-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/007,255
FILING DATE: 06-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/608,423
FILING DATE: 28-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/705,484
FILING DATE: 28-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/743,699
FILING DATE: 06-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Borucki, Andrea M
REGISTRATION NUMBER: 33651
REFERENCE/DOCKET NUMBER: 50301E
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-337-4847
TELEFAX: 317-337-4847
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 3132 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-10-242-056-60
Query Match 4.8%; Score 36.2; DB 9; Length 1714

```

Db 744 .G..M.T.H.TVBCSRD..SH..KB...T.DTHNCHT...T.HS..NABK...C.B..CS..683

QY 320 TTGCTACTTTTTCACCTGTGACGAGCTGTCCAACTTCTCTGTACTCTTTTGTCTCCCAAGC 379

Db 682 T...TAACHBHK.....AHSEYAG..YTDDB...NS.G..MH..CGC..M..D...TY..AS..DTN 623

QY 380 TTCCCGCCGCGCTCGAATGACAGATACAGAGACTCCCTCTGTGGACCCGTTGGAGAGTC 439

Db 622 ...TB...C.C..T.Y.Y..Y..A..SRS.ABB..TY..MHMBY.....KTHGGBYHD..DM..WBAC 563

QY 440 CAGAGACTTTTATTCACACTCACTTTTTCCTTTTTCATTTGACCTGGGCGGACGAGT 499

Db 562 ST..DM...NYS..ABY..B..CY..YHNATH..BRH..TBH..SSSS..BT..RS..GBY..HBBSD 503

QY 500 AAGTACTTATTTCTGTCAATCTGTGCAATCAAGATGCCCTGAGAGTGCACAGCCCCCTTC 559

Db 502 CCTGN...H..SY..T..SB..DSBD...YNY...BTAHT..SCB...H..R..NCHKH..SSS..SRC..443

QY 560 CCTCTTTTGGCGTCCGGAAGGGGTTCTTTTATGTCTTCACACCCCAACCTTCCCTC 619

Db 442 SM.CBT..C...B..S..MTBBC..SSCDR...EMCHCH...TCBT..H...CH..N..HC..BT..S..C..KS 383

QY 620 C 620

Db 382 H 382

## RESULT 2

```

US-10-184-634-120/c
Sequence 120. Application US/10184634
Publication No. US2003006868A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Collin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C217
CURRENT APPLICATION NUMBER: US/10/184,634
CURRENT FILING DATE: 2002-06-28
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 120
LENGTH: 1141
TYPE: PRT
ORGANISM: Homo Sapien
US-10-184-634-120

```

Query Match	5.4%	Score 40.6;	DB 9;	Length 1141;
-------------	------	-------------	-------	--------------

Best Local Similarity 11.0%; Prec.No. 0.011;  
Matches 53; Conservative 145; Mismatches 281; Indels 2; Gaps 1

Db 862 SC.AY.S.C.NSBAB.RMMSCBHY.S.KTY.CYBDCS.....TKCT.S.....BYT.803CS

QY 142 CCGTGGTCCCAAGGTGGCCCTCTCAACAGGAACCAATATTTTGTCTGTGACTGCT.201

QY 202 CTGAACAACAGCGGTGTGGGGGTGGGGAACAAGTGGATCTGGCCCTCTGTGAGGACACCTC.261

Db 802 YTSB.D.....S.AT.....B..AA..S.CSAS..R.AB.TCYKR.NC..G.BSTV.NS.743

QY 262 TGTGCTGCTGCGCCAGAGCTCTCTCTGTGGTGTCTCTCTGACCCGATG--CCTTTGACT.319

Db 742 .G..M.T.H.TYBCSD.SH..KB..T.DYHCHT...T.HS..NMBK...C.B..CS.663

[illegible]

### RESULT 3

```

US-10-123-155-144/C
? Sequence 144, Application US/10123155
? Publication No. US20030068794A1
? GENERAL INFORMATION:
? APPLICANT: Baker, Kevin P.
? APPLICANT: Beresini, Maureen
? APPLICANT: Deforge, Laura
? APPLICANT: Desnoyers, Luc
? APPLICANT: Filvaroff, Ellen
? APPLICANT: Gao, Wei-Qiang
? APPLICANT: Gerritsen, Mary E.
? APPLICANT: Goddard, Audrey
? APPLICANT: Godowski, Paul J.
? APPLICANT: Gueney, Austin L.
? APPLICANT: Sherwood, Steven
? APPLICANT: Smith, Victoria
? APPLICANT: Stewart, Timothy A.
? APPLICANT: Tumas, Daniel
? APPLICANT: Watanabe, Colin K
? APPLICANT: Wood, William
? APPLICANT: Zhang, Zemin
? TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
? FILE REFERENCE: P3330R10C30
? CURRENT APPLICATION NUMBER: US/10/123,155
? CURRENT FILING DATE: 2002-04-15
? Prior Application removed - See Palm or File Wrapper
? NUMBER OF SEQ ID NOS: 550
? SEQ ID NO 144
? LENGTH: 520
? TYPE: PRT
? ORGANISM: Homo Sapien
US-10-123-155-144

```

Query Match	Score	DB	Length
5.18;	38.6;	9;	520;
10.18;	38.6;	9;	520;
10.18;	38.6;	9;	520;

Matches 49; Conservative 135; Mismatches 301; Indels 0; Gaps 0;

```

OY 25 ACCGCTGGTGGTCCCTCGGCGCCAGAGGTCCTCGTGGTCTGTGACGACCATGCGCTTTG 318
      :::::  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 485 RYHHY...SCH.Y..TCSB...CCMH..SV.RR..B..CCYT.BRYTGM...XY.. 428
OY 317 ACTTGGTACTTTTCTACCTGAGCAGCGTGTCCAGTCTCTGACCTTTGTTGCTCCCA 378
      :::::  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 425 ...TTH.DV..CCTTG...CWB..M..CY.YNGYCCM..R.HCHN.SSTTT..N...CWB 366
OY 377 AGCTTCCCTCCGCGCCCTCGAGATCGAGATACGACGACTCCCTTGTGTGACCGCGTTTGAGA 436

```





## COMPUTER READABLE FORM:

```

MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Mb
COMPUTER: IBM compatible PC
OPERATING SYSTEM: Windows 95
SOFTWARE: Word 97 (DOS text format)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,584A
FILING DATE: 04-Feb-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Grady J. Frenchick
REGISTRATION NUMBER: 29,018
REFERENCE/DOCKET NUMBER: 16026,5180
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 257-3501
TELEFAX: (608) 257-2275
INFORMATION FOR SFO ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 444 bp
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Circular
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: no
IMMEDIATE SOURCE:
LIBRARY: plasmid, pGem3zf(+)
CLONE: G210
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 8p
US-09-018-584A-19

```

Query Match 4.2%; Score 31.4; DB 4; Length 444;

Best Local Similarity 64.4%; Pred. No. 2.2; Mismatches 26; Indels 0; Gaps 0;

```

Matches 47; Conservative 0;
QY 584 TTGCTTTATGCTTCACACCCACACCTTCCCTCCCTGCTGTTTGTGCCCA 643
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 335 TTCTTTCTCTCTCCCTCCCTCCCTCCACATCCCTCTCTCTCTCTCTCTTTGTTT 394
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 644 AAAACCAATGCT 656
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 395 AAAACCAATGCT 407

```

Search completed: July 8, 2003, 05:55:08  
 Job time : 198.615 secs

STATE: Wisconsin  
COUNTRY: U.S.A.  
ZIP: 53711-5399

US-RESULT 11  
US-08-921-195-1/c  
Sequence 1, Application US/08921195  
Patent No. 6147052  
GENERAL INFORMATION:  
APPLICANT: CHAU, RAYMOND M.W.  
TITLE OF INVENTION: ISOCATION AND USE OF ERYTHROID  
TITLE OF INVENTION: DIFFERENTIATION AND DEINCLEATION FACTORS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LYON & LYON  
STREET: 633 WEST SIXTH ST., SUITE 4700  
CITY: LOS ANGELES  
STATE: CA  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/921,195  
FILING DATE: 29-Aug-1997  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:

RESULT 12  
 US-08-785-420-1/c  
 Sequence 1, Application US/08785420  
 Patent No. 6001976  
 GENERAL INFORMATION:  
 APPLICANT: MacLennan, David H  
 APPLICANT: O'Brien, Peter J.  
 TITLE OF INVENTION: DIAGNOSIS FOR PORCINE MALIGNANT  
 HYPERTHERMIA  
 NUMBER OF SEQUENCES: 3  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Bell, Seltzer, Park & Gibson  
 STREET: P.O. Drawer 34009  
 CITY: Charlotte,  
 STATE: No. 6001976th Carolina 28234  
 COUNTRY: U.S.A.  
 ZIP: 28234  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/785,420  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/249,388  
 FILING DATE:  
 APPLICATION NUMBER: US 08/030,158  
 FILING DATE: 15-MAR-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Layton, Jr., Samuel G.  
 REGISTRATION NUMBER: 22807  
 REFERENCE/DOCKET NUMBER: 3477-73  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 704-377-1561

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/194,613  
FILING DATE: 30-NOV-1998  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Bucca Ph.D., Daniel  
REGISTRATION NUMBER: 42,368  
REFERENCE/DOCKET NUMBER: 5019-061  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-756-8000  
TELEFAX: 202-756-8087  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2478 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-09-194-613-1

Query Match 4.4%; Score 33.4; DB 4; Length 2478;  
Best Local Similarity 50.3%; Pred. No. 1.3;  
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 25 AAGAACTTCTTCTGAGAGCTTTCCACCCCTTCCCTGCTGAGCAGTGAAGT 84  
DB 2177 AAGGACTTGATGATGAGAGAGCTTCCCTCCGACTTCAGCAGGAGAGAGATGTGCG 2118  
QY 85 CAGGTAGGGAGACTCGGAGACTGCGATGTGCGCAGGAAAGGTGAGCGGGTGAAGTGCCT 144  
DB 2117 GCGGAGGAGCAAGAACCGCGCGCGGAGATTGTCACTGCGCGGTAGACACCTGTGAGACG 2058  
QY 145 GTTGCCAGGTGGCTCTTCAACAGAGAAACCAATATTTTT 187  
DB 2057 GGTGGCCAGGTGCTGTGCGGAGGAGAGATCTCTCATTTCTTG 2015

RESULT 9  
US-08-485-355B-51/c  
Sequence 51, Application US/08485355B  
Patent No. 6177075  
GENERAL INFORMATION:  
APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.  
TITLE OF INVENTION: Insect Viruses and Their Uses in  
Protecting Plants  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fleh, Hohbach Test Albritton & Herbert LLP  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,355B  
FILING DATE: 07-Jun-1995  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/440,522  
FILING DATE: 12-MAY-1995  
APPLICATION NUMBER: US 08/089,372  
FILING DATE: 08-JUL-1993  
APPLICATION NUMBER: AU P14081/92

FILING DATE: 14-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Treccartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2479 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 283..2307  
SEQUENCE DESCRIPTION: SEQ ID NO: 51:  
US-08-485-355B-51

Query Match 4.4%; Score 33.4; DB 4; Length 2479;  
Best Local Similarity 50.3%; Pred. No. 1.3;  
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 25 AAGAACTTCTTCTGAGAGCTTTCCACCCCTTCCCTGCTGAGCAGTGAAGT 84  
DB 2178 AAGGACTTGATGATGAGAGAGCTTCCCTCCGACTTCAGCAGGAGAGACATGTTGCG 2119  
QY 85 CAGGTAGGGAGACTGAGAGACTGCGATGTGCGCAGGAAAGGTGAGCGGGTGAAGTGCCT 144  
DB 2118 GCGGAGGAGCAAGAACGCGCGCGGAGATTGTCACTGCGCGGTAGACACCTGTGAGACG 2059  
QY 145 GTTGCCAGGTGGCTCTTCAACAGAGAAACCAATATTTTT 187  
DB 2058 GGTGGCCAGGTGCTGTGCGGAGGAGAGATCTCTCATTTCTTG 2016

RESULT 10  
US-09-072-596-323  
Sequence 323, Application US/09072596  
Patent No. 6458366  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Campos-Neto, Antonia  
APPLICANT: Houghton, Raymond  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Twardzik, Daniel R.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Hendrickson, Ronald C.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
TUBERCULOSIS  
NUMBER OF SEQUENCES: 350  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/072,596  
FILING DATE: 05-MAY-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.

ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,355B  
FILING DATE: 07-Jun-1995  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/440,522  
FILING DATE: 12-MAY-1995  
APPLICATION NUMBER: US 08/089,372  
FILING DATE: 08-JUL-1993  
APPLICATION NUMBER: AU PL4081/92  
FILING DATE: 14-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Treccartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2478 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 283..753  
SEQUENCE DESCRIPTION: SEQ ID NO: 47:  
US-08-485-355B-47  
Query Match 4.4%; Score 33.4; DB 4; Length 2478;  
Best Local Similarity 50.3%; Pred. No. 1.3;  
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;  
QY 25 AAGAACTTCTTCCTGGAGCCTTCCACCCCTTCCTGCTGAGCAGCGTAGTGG 84  
DB 2177 AACGACTGATGATGAGGAGCGTTGCTCGACTTCAGCAGGAGGACAGCATGTTGCG 2118  
QY 85 CAGGTAGGAGCTCGAGACTGCGATGATGTCAGAGAAAGGAGGAGCGGAGTGCCT 144  
DB 2117 GCGGAAGGACAGAAAGCGGCGCGGAGAGTTGTCAGTGCGGAGGAGTGTGACACCTGTGAGACG 2058  
QY 145 GTTGCCAGAGGTGGCTCTTCAACAGGAAACCAATATTTTGG 187  
DB 2057 GGTGGCCAGGTGCTGCGGAGGAGGAGATCTCCTCATTTCTTG 2015  
RESULT 7  
US-08-485-355B-49/c  
Sequence 49, Application US/08485355B  
Patent No. 6177075  
GENERAL INFORMATION:  
APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.  
TITLE OF INVENTION: Insect Viruses and Their Uses in  
Protecting Plants  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco

STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,355B  
FILING DATE: 07-Jun-1995  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/440,522  
FILING DATE: 12-MAY-1995  
APPLICATION NUMBER: US 08/089,372  
FILING DATE: 08-JUL-1993  
APPLICATION NUMBER: AU PL4081/92  
FILING DATE: 14-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Treccartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2478 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 366..2306  
SEQUENCE DESCRIPTION: SEQ ID NO: 49:  
US-08-485-355B-49  
Query Match 4.4%; Score 33.4; DB 4; Length 2478;  
Best Local Similarity 50.3%; Pred. No. 1.3;  
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;  
QY 25 AAGAACTTCTTCCTGGAGCCTTCCACCCCTTCCTGCTGAGCAGCGTAGTGG 84  
DB 2177 AACGACTGATGATGAGGAGCGTTGCTCGACTTCAGCAGGAGGACAGCATGTTGCG 2118  
QY 85 CAGGTAGGAGCTCGAGACTGCGATGATGTCAGAGAAAGGAGGAGCGGAGTGCCT 144  
DB 2117 GCGGAAGGACAGAAACGCGCGCGCGAGAGTTGTCAGTGCGGAGGAGTGTGACACCTGTGAGACG 2058  
QY 145 GTTGCCAGAGGTGGCTCTTCAACAGGAAACCAATATTTTGG 187  
DB 2057 GGTGGCCAGGTGCTGCGGAGGAGGAGATCTCCTCATTTCTTG 2015  
RESULT 8  
US-09-194-613-1/c  
Sequence 1, Application US/09194613  
Patent No. 6251654  
GENERAL INFORMATION:  
APPLICANT: GORDON, Karl H.  
APPLICANT: HANZLIK, Terry N.  
TITLE OF INVENTION: MODIFIED SMALL RNA VIRUSES  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McDermott, Will & Emery  
STREET: 600 13th Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3096

```
OY 401 GATACAGGACTCCCTTCTGTGACCCCTTTGGAGAGTCACAGACATTATATCAATCCAC 460
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1191 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1250
OY 461 TTTTTCCTTTTCATTTGGCCCGGGGCGGAGCGTAGTACTTATCTGTGATCC 520
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1251 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1310
OY 521 TGTCAACACAGCAAGTCCCTGAGTGCACAGCCCTTCCCTCTTTCGCTCGAAG 580
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1311 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1370
OY 581 GGTTCCTTTATGCTTCACACCCACCTTCCCTCCCTGCTGTTTGTGTC 640
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1371 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1430
OY 641 CCAAAAACAAGTCTCTCAACGTTT 667
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1431 YYYGTACCAATTTCTTATCTCTT 1457
```

## RESULT 4

```
US-09-007-005-17/c
; Sequence 17, Application US/09007005B
; Patent No. 628358
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihne
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; EARLIER FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17
```

Query Match Best Local Similarity 4.9%; Score 37.2; DB 4; Length 289;

Matches 36; Conservative 97; Mismatches 149; Indels 0; Gaps 0;

```
OY 180 TATTTTGTCTGACTGCTCTGAAACAGGCGCTGGGGGTGGGAGACACTTGA 239
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 284 TTTTTCCTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTT 225
OY 240 TGTGCGCTCTGAGACACCTCTGCTGCTGCTGCGCCAGGCTCCTGCTGCTTCT 299
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 224 CTYGGSYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNY 165
OY 300 CTGAGCCGATGCTTGTGACTTTTCTACTTTTCTAGACAGTCTCCAGTCTCTGC 359
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 SYNYNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNY 105
OY 360 TACCTTTTTCCTGCTGAGGCTTCCCTGCGCTCCAGTCAAGAGATACAGGACTCCCTCT 419
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 104 SYNYNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNY 45
```

```
OY 420 GTGAGCCGTTTGGAGAGTCCAGACACTTATATCAATCCACT 461
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 44 AYAATYTYGYTAAVAATYAAVGYTAAVAATYTYGYTYCY 3
```

## RESULT 5

```
US-09-244-796-17/c
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihne
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; EARLIER FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17
```

Query Match Best Local Similarity 4.9%; Score 37.2; DB 4; Length 289;

Matches 36; Conservative 97; Mismatches 149; Indels 0; Gaps 0;

```
OY 180 TATTTTGTCTGACTGCTCTGAAACAGGCGCTGGGGGTGGGAGACACTTGA 239
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 284 TTTTTCCTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTT 225
OY 240 TGTGCGCTCTGAGACACCTCTGCTGCTGCTGCGCCAGGCTCCTGCTGCTTCT 299
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 224 CTYGGSYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNYNNY 165
OY 300 CTGAGCCGATGCTTGTGACTTTTCTACTTTTCTAGACAGTCTCCAGTCTCTGC 359
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 SYNYNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNY 105
OY 360 TACCTTTTTCCTGCTGAGGCTTCCCTGCGCTCCAGTCAAGAGATACAGGACTCCCTCT 419
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 104 SYNYNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNYSYNNY 45
```

## RESULT 6

```
US-08-485-355B-47/c
; Sequence 47, Application US/08485355B
; Patent No. 617075
; GENERAL INFORMATION:
; APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.
; TITLE OF INVENTION: Insect Viruses and Their Uses in
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
```

TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2887 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-983-502-14

Query Match 16.4%; Score 123.6; DB 4; Length 2887;  
Best Local Similarity 93.5%; Pred. No. 2.4e-29;  
Matches 129; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 13 TATTGAAGTAAAGAACTTCTCTCTGAGAGCCTTCCACCCCTTCCCTGCTGAGCA 72  
DB 144 TATTGAAGTAAAGAACTTCTCTCTGAGAGCCTTCCACCCCTTCCCTGCTGAGCA 203  
QY 73 CGTGAGTTAGGACGAGTTAGGGGACTCGAGACTGCGATGTCGCCAGAAAGGTGAGC 132  
DB 204 CGTGAGTTAGGACGAGTTAGGGGACTCGAGACTGCGATGTCGCCAGAAAGGTGAGC 263  
QY 133 GGGTGAGTGCCTGTGCC 150  
DB 264 GGATTATATCTCTCTGCC 281

## RESULT 2

PCT-US96-10521-14  
Sequence 14, Application PC/TUS9610521  
GENERAL INFORMATION:

APPLICANT:  
TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS RECEPTORS  
NUMBER OF SEQUENCES: 34  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30 (EPC)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10521  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: IL 114,615  
FILING DATE: 16-JUL-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: IL 114,986  
FILING DATE: 17-AUG-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: IL 115,319  
FILING DATE: 14-SEP-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: IL 116,588  
FILING DATE: 27-DEC-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: IL 117,932  
FILING DATE: 16-APR-1996  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2887 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
PCT-US96-10521-14

Query Match 16.4%; Score 123.6; DB 5; Length 2887;  
Best Local Similarity 93.5%; Pred. No. 2.4e-29;  
Matches 129; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 13 TATTGAAGTAAAGAACTTCTCTCTGAGAGCCTTCCACCCCTTCCCTGCTGAGCA 72  
DB 144 TATTGAAGTAAAGAACTTCTCTCTGAGAGCCTTCCACCCCTTCCCTGCTGAGCA 203  
QY 73 CGTGAGTTAGGACGAGTTAGGGGACTCGAGACTGCGATGTCGCCAGAAAGGTGAGC 132  
DB 204 CGTGAGTTAGGACGAGTTAGGGGACTCGAGACTGCGATGTCGCCAGAAAGGTGAGC 263  
QY 133 GGGTGAGTGCCTGTGCC 150  
DB 264 GGATTATATCTCTCTGCC 281

## RESULT 3

US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:

APPLICANT: DORNER, F.  
APPLICANT: SCHETFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-Fls  
US-08-232-463-14

Query Match 11.1%; Score 83.8; DB 1; Length 7218;  
Best Local Similarity 3.4%; Pred. No. 2e-16; Mismatches 128; Indels 0; Gaps 0;  
Matches 13; Conservative 246; Mismatches 128; Indels 0; Gaps 0;

QY 281 TCTCTGTTGTTCTCTCTGAGCGAGTTCGACTTGTGACTTTTCACTGAG 340  
DB 1071 YY 1130  
QY 341 CAGTCTCAGTCTCTCTGCTACCTTTTGTCTCCAAAGCTTCCCTGCGCTGAGTGA 400  
DB 1131 YY 1190

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:45:20 ; Search time 180.615 Seconds  
(without alignments)  
1278.561 Million cell updates/sec

Title: US-09-477-082-2  
Perfect score: 753  
Sequence: 1 aatagaccgcgcgtatgtgaaa.....taactcgtttttaacctt 753

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapept 1.0  
Searched: 441362 segs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/prodata/1/lna/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/1/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/lna/5A\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/lna/5B\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/lna/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/lna/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	123.6	16.4	2887	US-08-883-502-14	Sequence 14, Appl
2	123.6	16.4	2887	PCT-US96-10521-14	Sequence 14, Appl
3	83.8	11.1	7218	US-08-332-463-14	Sequence 14, Appl
4	37.2	4.9	289	US-09-007-005-17	Sequence 17, Appl
5	37.2	4.9	289	US-09-244-796-17	Sequence 17, Appl
6	33.4	4.4	2478	US-08-485-355-47	Sequence 47, Appl
7	33.4	4.4	2478	US-08-485-355-49	Sequence 49, Appl
8	33.4	4.4	2478	US-09-194-613-1	Sequence 1, Appl
9	33.4	4.4	2478	US-08-485-355B-51	Sequence 51, Appl
10	32.8	4.3	1166	US-09-072-596-323	Sequence 323, App
11	32.4	4.3	2721	US-08-921-195-1	Sequence 1, Appl
12	32.4	4.3	15378	US-08-885-420-1	Sequence 1, Appl
13	31.8	4.2	277	US-09-007-005-3	Sequence 3, Appl
14	31.8	4.2	277	US-09-244-796-3	Sequence 3, Appl
15	31.4	4.2	444	US-09-018-584A-19	Sequence 19, Appl
16	31.4	4.2	444	US-09-018-584A-13	Sequence 13, Appl
17	31	4.1	2088	US-09-351-414-3	Sequence 3, Appl
18	31	4.1	4415	US-09-486-580A-1	Sequence 1, Appl
19	31	4.1	99500	US-09-798-096-10	Sequence 10, Appl
20	30.8	4.1	4978	US-08-220-603A-1	Sequence 1, Appl
21	30.8	4.1	176373	US-09-128-155-17	Sequence 17, Appl
22	30.6	4.1	867	US-09-071-035-221	Sequence 221, App
23	30.6	4.1	8321	US-08-680-506-1	Sequence 1, Appl
24	30.6	4.1	12368	US-08-976-259-20	Sequence 20, Appl
25	30.4	4.0	248	US-09-007-005-32	Sequence 32, Appl
26	30.4	4.0	248	US-09-244-796-32	Sequence 32, Appl
27	30.4	4.0	909	US-09-247-135-131	Sequence 131, App

28	30.4	4.0	17410	1	US-07-841-646-3	Sequence 3, Appl
29	30.4	4.0	17410	1	US-08-147-023-3	Sequence 3, Appl
30	30.4	4.0	17410	1	US-08-447-570-3	Sequence 3, Appl
31	30.4	4.0	17410	2	US-08-449-700-3	Sequence 3, Appl
32	30.4	4.0	17410	2	US-08-449-699A-3	Sequence 3, Appl
33	30.4	4.0	17415	3	US-08-486-343A-1	Sequence 1, Appl
34	30.4	4.0	17415	5	PCT-US95-07339-1	Sequence 1, Appl
35	30.4	4.0	152331	3	US-09-128-135-16	Sequence 16, Appl
36	30.2	4.0	840	1	US-08-690-457-14	Sequence 14, Appl
37	30.2	4.0	840	2	US-08-628-187-14	Sequence 14, Appl
38	30.2	4.0	1702	3	US-08-132-649-3	Sequence 3, Appl
39	30.2	4.0	1702	3	US-08-767-579-3	Sequence 3, Appl
40	30.2	4.0	2080	1	US-08-132-648-1	Sequence 1, Appl
41	30.2	4.0	2080	3	US-08-767-579-1	Sequence 1, Appl
42	30.2	4.0	98844	4	US-09-791-211-10	Sequence 10, Appl
43	29.8	4.0	118	3	US-08-833-167-46	Sequence 46, Appl
44	29.8	4.0	118	4	US-09-344-837A-46	Sequence 46, Appl
45	29.8	4.0	4699	1	US-08-487-890A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-08-983-502-14  
Sequence 14, Application US/08983502  
Patent No. 6399327  
GENERAL INFORMATION:  
APPLICANT: David WALLACH  
APPLICANT: Mark P. BOLDIN  
APPLICANT: Tanaya M. GONCHAROV  
APPLICANT: Yuri V. GOLITSEV  
TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS RECEPTORS  
TITLE OF INVENTION: AND OTHER PROTEINS  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Broadway and Neimark  
SUITE: 419 Seventh Street N.W., Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/983,502  
FILING DATE: 16-JUN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10521  
FILING DATE: 14-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 114,615  
FILING DATE: 16-JUL-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 114,986  
FILING DATE: 17-AUG-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 115,319  
FILING DATE: 14-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 116,588  
FILING DATE: 27-DEC-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 117,932  
FILING DATE: 16-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Broadway, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: WALLACH=19  
TELECOMMUNICATION INFORMATION:



Fri Jul 11 09:23:58 2003

us-09-477-082-2.rst

Page 9

[illegible]

```

search completed: July  8, 2003, 05:51:12
Job time : 5682.6 secs

```



[illegible]

	Query Match	5.7%:	Score 42.8:	DB 17:	Length 1101:	
	Best Local Similarity	48.7%:	Pred. No. 2:			
	Matches	Conservative	19:	Mismatches	40:	Indels
						Gaps 0:
OY	552 CCCCTTCCCTTCCTTGCGCGTCTGAAGGGTTTCATTATGCTTCACACCACCCTCCT	611				
Dd	987 MCTMTTGTTHTCMCTGCCCAATAATTTCTTCHTTTTTYYVCCCCSSCCCCC	1046				
OY	612 TTCCCCTCCCTCTGTTTGTGTGCCAAACAACACTTCTTAACGTTT	666				
Dd	1047 CCCCCCCCCCCCCCTTTTMTTMMWAAAAMTWAMTWMAAATTTT	1101				
	RESULT 13					
	CNS016FT					
	LOCUS					
	DEFINITION	CNS016FT	1201 bp	DNA	Linear	GSS 26-JUL-1999
	ACCESSION	Drosophila melanogaster genome survey sequence Sp6 end of BAC				
	VERSION	BACN15102 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
	KEYWORDS	AI106691				
SOURCE	GI:5623268					
ORGANISM	GSS.					
	Drosophila melanogaster.					
	Drosophila melanogaster.					
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;					
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;					
	Ephydroidea; Drosophilidae; Drosophila.					
	1 (bases 1 to 1201)					
REFERENCE	Genoscope.					
AUTHORS	Direct Submission					
TITLE	Submitted (22-JUL-1999) Genoscope - Centre National de Sequencage :					
JOURNAL	Bp 191 91006 ENRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr					
	- Web : www.genoscope.cns.fr)					
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (www.edgp.ac.uk) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelosBAC11.					
FEATURES	Location/Oualifiers					
SOURCE	1..1201					
	/organism="Drosophila melanogaster"					
	/db_xref="taxon:7227"					
	/clone_1ld="DrosBAC"					
	/plasmid="pBelosBAC11"					
	/note="end : SP6"					
BASE COUNT	272 a 207 c 106 g 438 t 178 others					
ORIGIN						
	Query Match	5.7%:	Score 42.8:	DB 17:	Length 1201:	
	Best Local Similarity	33.1%:	Pred. No. 2:			
	Matches	Conservative	54:	Mismatches	124:	Indels
						Gaps 0:
OY	461 TTTTTCCTTCATTCATTTGCCTGGGGGCGGACGGTAGTACTTATTCGTGCATTC	520				
Dd	782 TTTTTCCTTCATTCATTTMMKMMAMNMCMMACTAACMAAATCATMTTKATCTCTA	841				
OY	521 TGTCGATCACGAATGCCCTGAGGTGCACAGCCCTTTCCCTCTTTCCGCTCGTAAGG	580				
Dd	842 AATAAAMHMRCCTTACTTACTTACCTTACCCHCCMCYCCTTACCACYCCMYCCYCH	901				
OY	581 GGTTCCTCTATFAGTCCTCACCCGACACCTTCCTCCCTCGCTGCGCCTGTGTTGTGC	640				
Dd	902 YTTTTCCTTCATTCATTTTTHATTTTTHATTTTTHATTTTTHATTTTTHATTTT	961				
OY	641 CCAAAAAACAAGTCTCTAAAGCTTTTCGATGTGGAATTCGGGAAATTAACCTGCACCC	700				
Dd	962 ATTMCAAHMCMCYTTTMMWTATTTTYTMMTCMWYCHTTHHTTTTYYVCYTATAT	1021				

BASE COUNT	229 a	/clone_lib="RRC1-98"	357 t
ORIGIN	276 c	/note="end : T7"	150 others
	89 g		

Query Match	5.9%;	Score 44.8;	DB 17;	Length 1101;
Best Local Similarity	32.1%;	Pred. No. 0.61;		
Matches 117;	Conservative 66;	Mismatches 182;	Indels 0;	Gaps 0;

```
QY      265 TGCATGCCCTGGCCCAAGCTCTCCTGTTGTCTCTGAGCGCATGACCTTGTCTGTCT 324
          ::::: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      652 KBBBBSKTTTGAKSKAYTTWTATTATTTTTTCCTTTTTTTTTTTCTTTTTT 711
```

325 ACCTTTTCCTCTGACGACAGTCTCCAGTTCCTGCTACCTTTTGTCCTCCAGCTATCCC 384  
 |||| | | | | : : : : : : : : : : : : : : : :  
 Db 712 TTCTTTTCCTCTTTTTTTCCTCCCTCCCTTTTTCCTCTCTCTCTCTCTCTCTCTCTCTCT 771

581 TACCCTCCGAGATGCTGATACGAGATCCCTCTCTGAGCCGCTTGGAGAGTCCGAG 44  
 ||||:| :|||:| :|||  
 Db 772 CCGCCCGCYCTCCGCCCTCTTTTTCYCYCTTTCGYYNCTCGCCCTTTTTCYGCT 831

[illegible]

```

      ||| | | :: : | | :: | | ::||| |
Db    892 CCTTTCTCTTCYCYYYCYTYTCCGCCGCCYCCGCCGCCYCTTTT 951
      ||| | | :: : | | :: | | ::||| |
E E C TTTCCCCCAGGCAACCGCCCTTCCCCTTTATCATGGGGAACGCGGCAGG

```

Dv	625 CCGCT 629
Db	952 TTTCCTTTTNTTCTTTTTCCTTTTTCCTTTTTCCTTTTCCTCCKCCYTCGCCCYCCCTCTCATYC 101

Db 1012 TCCYT 1016

AL514103/c	AL514103	597 bp	mRNA	linear	EST 13-FEB-2000
LOCUS					
DEFINITION	AL514103	L71_NFL006-PL2	Homo sapiens	CDNA clone	CL08A0052F03 3

ACCESSION AL514103  
VERSION AL514103.1 GI:12777597  
KEYWORDS EST.

ORGANISM	Homio sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	

**AUTHORS** Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
**TITLE** Full-length cDNA libraries and normalization  
**JOURNAL** Unpublished (2001)  
**COMMENT** Contact: Genoscreen

**FEATURES**  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: [segrete@genoscope.cns.fr](mailto:segrete@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr).  
Location/Qualifiers

```

source
1. .39/
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="C10BA0052F03"

```

```

/vecvec_110- br_1n1000_02
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6, Site_1: NotI, 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end

```

cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

BASE COUNT	102 a	70 c	231 g	112 t	82 others
ORIGIN	Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"				

Query Match	5.9%	Score	44.4	DB	9	length	597
Best Local Similarity	41.5%	Pred	0.78				
Matches	78	Conservative	28	Mismatches		Indels	0
						Gaps	0

Oy 457 CCACTTTTTCATTTGACATTGGCCGTGGAGCGAAGACTTTATCTGC 516  
|| : ||| | ||| : ||| :  
Db 330 CCCYVTTTTTTTITTTGGGGGCCGGGYTTTTCCCCTTTTITTMCCCCCCT 271

GY 517 AATTCGTGGAATACGAATGCCCTGAGCGTCAAGACGGCCCTTTCCCTCTTCGCCTCG 576  
       ::                ::                ::                ::                :  
 Db 270 TATTATTTTTTCCCCCCCYCATTTCYCCSSSCCCYCCCCCBGTTTTKGGCCCY 211

CY	377	AAGSSAIIICCTIITATGTCITACCCCTCCTTTCCTCCCTCTGTATTTAS	638
DB	210	YKGGGRKTITTTTTTKSGGKKCCCCCCCCCCCCCCCXYHMTTTTTTTTTTTT	151

DB 150 TTTCMMV 143

LOCUS	910 bp	DNA	linear	GSS 03-JUN-1999
CNS0060N				
CNS0060R				
RESULT 11				

ACCESSION  
NUMBER  
BACR44521 of RCP1-98 library from *Drosophila melanogaster* (fruit  
fly); genomic survey sequence.  
AL065629  
AL065630 1 CT-0044508  
INSTRUMENT

KEYWORDS  
GSS.  
SOURCE  
ORGANISM  
*Drosophila melanogaster*.  
*Drosophila melanogaster*  
*Eukaryota*.  
*Metazoa*. *Arthropoda*. *Neurynoda*. *Insecta*. *Dipterovgota*.

REFERENCE  
1 (bases 1 to 910)  
Genoscope.

Received: 02-JUN-1999  
 Submitted to: Genoscope - Centre National de Séquençage  
 BP 191 91006 EVRY cedex - FRANCE (E-mail: [segrat@genoscope.cns.fr](mailto:segrat@genoscope.cns.fr))  
 - web: [www.genoscope.cns.fr](http://www.genoscope.cns.fr)

collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information

melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mammone in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

ECORI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual cDNA clones, the entire library or

FEATURES  
source 1. 910  
filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
location/Qualifiers

```

/organism="Drosophila
/db_xref="taxon:7227
/clone="BACR14J21"
/clone_1b="RPC1-98"
metanogaster

```

BASE COUNT ORIGIN	202 a	63 c	112 g	198 t	335 others
----------------------	-------	------	-------	-------	------------

Best Local Similarity 18.6%; Pred. No. 16;

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN37J10"
/clone_lib="DrosBAC"
/plasmid="PbeloBAC11"
/notes="end : Sp6"

BASE COUNT      258 a      174 c      277 g      120 t      272 others
ORIGIN

Query Match      6.2%; Score 46.4; DB 17; Length 1101;
Best Local Similarity 26.0%; Pred. No. 0.24;
Matches 95; Conservative 95; Mismatches 176; Indels 0; Gaps 0;

QY 270 CCTGGCCAGGCTCTCTGTGTCTCTGTCTGTGACCGATGCTTGTACTTGT 329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1085 CCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1026
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 330 TTCACCTCTAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1025 TTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 966
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 390 CCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 965 TTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 906
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 450 TATCAATCACTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 509
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 905 TTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 846
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 510 TTTCTGATCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 569
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 845 TTTCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT 786
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 570 GGTCTGAAGGGGTTCTTTATGTTCTTCACACCCCAACCTTTCCCTG 629
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 785 CTTCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT 726
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 630 GTTTT 635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 725 TCTTTT 720

RESULT 8
CNS0181N/C      1101 bp      DNA      linear      GSS 26-JUL-1999
LOCUS      Drosophila melanogaster genome survey sequence Sp6 end of BAC
DEFINITION      BACN37J10 of DrosBAC library from Drosophila melanogaster (fruit
                fly), genomic survey sequence.
ACCESSION      AL108773
                GI:5629077
VERSION      GSS.
KEYWORDS      Drosophila melanogaster.
SOURCE      Drosophila melanogaster.
ORGANISM      Drosophila melanogaster.
                Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
                Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
                Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 1101)
                Genoscope.
                Direct Submission
                Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
                BP 191 91006 EVRY cedex - FRANCE (E-mail : seget@genoscope.cns.fr
                - Web : www.genoscope.cns.fr)
                Determination of this BAC-end sequence was carried out as part of a
                collaboration with the European Drosophila Genome Project (BDGP) -
                http://www.edgp.ebl.ac.uk - This Drosophila melanogaster BAC
                library (Dros BAC) was made by Alain Billaud at CDPH (Centre
                d'Etude du Polymorphisme Humain) with funding provided by a MRC
                project grant. The DNA was prepared from embryos by Alain Bucheton
                and Genevieve Payan. It has been constructed in the vector
                PbeloBAC11.

FEATURES
Source      1..1101
                Location/Qualifiers
                /organism="Drosophila melanogaster"

```

```

/db_xref="taxon:7227"
/clone="BACN37J10"
/clone_lib="DrosBAC"
/plasmid="PbeloBAC11"
/notes="end : Sp6"

BASE COUNT      125 a      364 c      197 g      116 t      299 others
ORIGIN

Query Match      6.1%; Score 46.2; DB 17; Length 1101;
Best Local Similarity 21.5%; Pred. No. 0.27;
Matches 59; Conservative 99; Mismatches 117; Indels 0; Gaps 0;

QY 452 TCAATCCACTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 511
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1028 YCCCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 969
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 512 CTGCTATCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 571
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 968 CCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 909
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 572 TCTGAAGGGGTTCTTTATGTTCTTCACACCCCAACCTTTCCCTGCTGT 631
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 908 CTTCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT 849
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 632 TTTGTTGCCCAAAACAGTTCTTAAGTTTGTGATGATGATGATGATGAT 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 848 TTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 789
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 692 CTTGACCCCTTTGCAATGATGATGATGATGATGATGATGATGATGAT 726
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 788 CTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 754

RESULT 9
CNS00L6Z      1101 bp      DNA      linear      GSS 03-JUN-1999
LOCUS      Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION      BACR24F06 of Rpci-98 library from Drosophila melanogaster (fruit
                fly), genomic survey sequence.
ACCESSION      AL068148
                GI:4958076
VERSION      GSS.
KEYWORDS      Drosophila melanogaster.
SOURCE      Drosophila melanogaster.
ORGANISM      Drosophila melanogaster.
                Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
                Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
                Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 1101)
                Genoscope.
                Direct Submission
                Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
                BP 191 91006 EVRY cedex - FRANCE (E-mail : seget@genoscope.cns.fr
                - Web : www.genoscope.cns.fr)
                Determination of this BAC-end sequence was carried out as part of a
                collaboration with the Berkeley Drosophila Genome Project (BDGP).
                The BDGP is constructing a physical map of the Drosophila
                melanogaster genome using these BACs. For further information
                please see http://www.fruitfly.org the BDGP Drosophila
                melanogaster BAC library was prepared by Kazutoyo Osogawa and
                Aaron Mammeter in Pieter de Jong's laboratory in the Department of
                Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
                NY. The library is named Rpci-98 and was constructed by partial
                EcoRI digestion of Drosophila DNA provided by the BDGP from the
                isogenic strain y2; cn bw sp. the same strain used for the BDGP's
                P1 and EST libraries. A more detailed description of the library
                and how to order individual BAC clones, the entire library, or
                filters for hybridization from the BACPAC Resource Center can be
                found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
Source      1..1101
                Location/Qualifiers
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
                /clone="BACR24F06"

```



[illegible][illegible]

FEATURES		Location/Qualifiers	
SOURCE		1..522	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="DKFZp613f1234"	
		/clone_id="313 (synonym: hlcc2)"	
		/dev_stage="adult"	
		/lab_host="DH10B"	
		/note="vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB; cDNA-collection"	
BASE COUNT		120 a	116 c 133 g 149 t 4 others
ORIGIN			
Query Match	16.6%; Score 125.2; DB 9; Length 522;		
Best local similarity	94.2%; Pred. No. 1.4e-21;		
Matches 130; Conservative	0; Mismatches 8; Indels 0; Gaps 0;		
QY	13 TATTGAAAGTAAAGAAACTTTCTTCCTGGAGACCTTCCACCCCTCTCCCTGCTGAGCA 72		
Db	226 TATTGAAAGTAAAGAAACTTTCTTCCTGGAGACCTTCCACCCCTCTCCCTGCTGAGCA 285		
QY	73 CGTGAAGTTTGGCGAGTTAGGGGACCTGGAGACCTGCATGGTGCAGAGAAAGGCTGAGGC 132		
Db	286 CGTGAAGTTTGGCGAGTTAGGGGACCTGGAGACCTGCATGGTGCAGAGAAAGGCTGAGGC 345		
QY	133 GGGTAGTGTCCCTGTGCC 150		
Db	346 GGATTATATTCTCTCTGCC 363		
RESULT 2			
LOCUS	AL702116 421 bp mRNA linear EST 22-MAR-2002		
DEFINITION	DKFZp686E18155_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone		
ACCESSION	AL702116		
VERSION	AL702116.1 GI:19685471		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 421)		
AUTHORS	Bloecher,H., Boecher,M., Brandt,P., Mewes,W., Well,B. and Wiemann ,S.		
TITLE	EST (Bloecher,H., Boecher,M., Brandt,P., Mewes,H.W., Well,B. and Wiemann,S.)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Bloecher H MIPS		
	Am Klopferspitz 18a D-82152 Martinsried, Germany		
	This is the 5' sequence of the clone insert		
	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer		
	Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;		
	financed by GSF (National Research Centre for Biotechnology Ltd.,		
	Brunschweig/Germany) within the cDNA sequencing consortium of the		
	German Genome Project.		
	No 81 sequence available.		
	This clone (DKFZp686E18155) is available at the RZPD in Berlin.		
	Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059		
	Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.		
FEATURES			
SOURCE			
	1..421		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="DKFZp686E18155"		
	/clone_id="686 (synonym: hlcc3)"		
	/tissue_type="human skeletal muscle"		
	/dev_stage="adult"		
	/lab_host="DH10B"		
	/note="vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB; cDNA-collection"		
BASE COUNT		95 a	89 c 113 g 124 t

ORIGIN	Query Match	16.4%	Score 123.6	DB 9	Length 421
Best Local Similarity	93.5%	Pred. No. 3,7e-21			
Matches 129	Conservative 0	Mismatches 9	Indels 0	Gaps 0	
QY	13 TATTGAAGTAAAGAAACCTCTTCCTGGGAGCCTTCCACCCCTTCCTGCTGAGCA	72			
Db	207 TATTGAAGTAAAGAAACCTCTTCCTGGGAGCCTTCCACCCCTTCCTGCTGAGCA	266			
QY	73 CGTGGAGTTAGGACAGCTTGGGGAGCTGGAGACTGCGATGGTGGCAGGAAAGGTGGAGC	132			
Db	267 CGTGGAGTTAGGACAGCTTGGGGAGCTGGAGACTGCGATGGTGGCAGGAAAGGTGGAGC	326			
QY	133 GGGTGAAGTCCCTGTGGC	150			
Db	327 GGATTATATTTCTCTGCC	344			
RESULT 3					
BI824550	804 bp	linear	EST 04-OCT-2001		
LOCUS	60303515661 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5176548 5'				
DEFINITION	nrna sequence.				
ACCESSION	BI824550				
VERSION	BI824550.1	GI:15936100			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
TITLE	1 (bases 1 to 804)				
JOURNAL	NIH-MGC http://mgc.scl.nih.gov/.				
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)				
	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgapbs-remail.nih.gov				
	Tissue Procurement: Life Technologies, Inc.				
	cDNA Library Preparation: Life Technologies, Inc.				
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
	DNA Sequencing by: Incyte Genomics, Inc.				
	Clone distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LLNL at:				
	http://image.llnl.gov				
	plate: LLM11439	row: 0	column: 13		
	High quality sequence stop: 620.				
FEATURES	Location/Qualifiers				
source	1..804				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone_image="5176548"				
	/clone_id="NIH_MGC_115"				
	/lab_host="DH10B"				
	/note="Organ: pooled brain, lung, testis; Vector: pCMV-Sport6; site:1: NotI; site:2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is 0.190-dt primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC Library."				
BASE COUNT	226 a	187 c	209 g	182 t	
ORIGIN					
Query Match	12.8%	Score 96.2	DB 13	Length 804	
Best Local Similarity	85.1%	Pred. No. 4e-14			
Matches 120	Conservative 0	Mismatches 18	Indels 3	Gaps 1	
QY	13 TATTGAAGTAAAGAAACCTCTTCCTGGGAGCCTTCCACCCCTTCCTGCTGAGCA	72			
Db	202 TATTGAAGTAAAGAAACCTCTTCCTGGGAGCCTTCCACCCCTTCCTGCTGAGCA	261			



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:04:24 ; Search time 5677.6 Seconds  
(without alignments)  
2147.949 Million cell updates/sec

Title: US-09-477-082-2

Perfect score: 753  
Sequence: 1 aatgaccgcgtattgaa.....taccctgttttaacctt 753

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST :  
1: em\_estba : \*  
2: em\_estchum : \*  
3: em\_estcin : \*  
4: em\_estlmu : \*  
5: em\_estlov : \*  
6: em\_estplo : \*  
7: em\_estro : \*  
8: em\_hic : \*  
9: gb\_estcl : \*  
10: gb\_estl2 : \*  
11: gb\_hic : \*  
12: gb\_est3 : \*  
13: gb\_est4 : \*  
14: gb\_est5 : \*  
15: em\_estfun : \*  
16: em\_estom : \*  
17: gb\_gss : \*  
18: em\_gss\_hum : \*  
19: em\_gss\_inv : \*  
20: em\_gss\_pln : \*  
21: em\_gss\_vrt : \*  
22: em\_gss\_fun : \*  
23: em\_gss\_mam : \*  
24: em\_gss\_mus : \*  
25: em\_gss\_other : \*  
26: em\_gss\_pro : \*  
27: em\_gss\_rod : \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	125.2	16.6	522	9	AL600571
2	123.6	16.4	421	9	AL702115
3	96.2	12.8	804	13	BT824550
4	58.6	7.8	987	17	CNS0057E
5	55.4	7.4	987	17	CNS00418
6	46.8	6.2	772	17	CNS02BMV

Result No.	Score	Query Match	Length DB	ID	Description
7	46.4	6.2	1101	17	CNS017RP
8	46.2	6.1	1101	17	CNS0181N
9	44.8	5.9	1101	17	CNS0016Z
10	44.4	5.5	597	9	AL514103
11	43.2	5.7	910	17	CNS0060N
12	42.8	5.7	1101	17	CNS0160T
13	42.8	5.7	1201	17	CNS016FT
14	42.4	5.6	571	17	AQ646132
15	42.2	5.6	846	17	CNS00KOY
16	42.2	5.6	1101	17	CNS00Z4D
17	42.2	5.6	1143	17	B11483
18	41.8	5.5	1124	17	B11278
19	41.2	5.5	1100	17	CNS00FSE
20	41.2	5.4	362	9	A1349834
21	41.1	5.4	739	17	BH179653
22	41.1	5.4	739	17	CNS0710R
23	41.1	5.4	1100	17	CNS00DEX
24	41.1	5.4	1107	17	AG129299
25	40.8	5.4	262	13	BM119154
26	40.8	5.4	1061	17	CNS001FJ
27	40.8	5.4	1201	17	CNS0160T
28	40.6	5.4	914	17	AG184648
29	40.4	5.4	980	17	CNS03844
30	40.2	5.3	934	17	CNS00D4F
31	40.2	5.3	1117	17	CNS00D2S
32	40.2	5.3	625	10	AM053904
33	40.0	5.3	1101	17	CNS00GC9
34	40.0	5.3	1128	13	B1956472
35	39.8	5.3	439	17	BH749911
36	39.8	5.3	1101	17	CNS0030C
37	39.8	5.3	1220	14	BQ058373
38	39.8	5.3	1352	12	BE896162
39	39.8	5.3	1364	13	BM474596
40	39.6	5.3	377	17	AO820430
41	39.6	5.3	1201	17	CNS016F2
42	39.4	5.2	693	17	AG043041
43	38.4	5.2	919	17	CNS006S5
44	38.4	5.2	1201	17	CNS0010T
45	39.2	5.2	680	17	AG102143

## ALIGNMENTS

RESULT 1  
LOCUS AL600571 522 bp mRNA linear EST 14-AUG-2001  
DEFINITION DKF2P313F1234\_r1 313 (synonym: hicc2) Homo sapiens CDNA clone  
ACCESSION DKF2P313F1234.5, mRNA sequence.  
VERSION AL600571  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 522)  
Ansoerge,W., Winkner,U., Mewes,W., Weil,B. and Wiemann,S.  
EST (Ansoerge,W., Winkner,U., Mewes,H.W., Weil,B. and Wiemann,S.)  
Unpublished (1999)  
CONTACT: Ansoerge W  
MIPS

Am Klopferst 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
Sequenced by EMBL (European Molecular Biology Laboratories,  
Heidelberg/Germany) within the CDNA sequencing consortium of the  
German Genome Project.  
No 5' sequence available.  
This clone (DKF2P313F1234) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

DR P-PSDB; ABB59675.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from *Drosophila* and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 5816; 21pp + Sequence Listing: English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 XX  
 SO Sequence 7723 BP; 2272 A; 1661 C; 1868 G; 1902 T; 0 other:  
 Query Match 4.9%; Score 37.2; DB 23; Length 7723;  
 Best Local Similarity 49.5%; Pred. No. 1.9;  
 Matches 96; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
 QY 280 GTCCTCCGTGCTGCTCTCTCTGAGCGGATGCTTGAAGCTTGTACTTTTTCAGCTCTGA 339  
 DB 2688 GCGCTTCTTCCCGCTTTTGTTCCTCACTCAATAGCTTGCGCTTTTTCGCTTA 2629  
 QY 340 GCACTCTCAGTTCCTCTGCTACCTTTTGTCTCCAGCTTCCGCGCTCGAATGC 399  
 DB 2628 ACCGCTCCATCTCTCTGCGCAGAGATTAGATCAGACTTGCGCTTGTGACTTT 2569  
 QY 400 AGATACAGGACGCCCTCTGTGTGAGCCGTTTGAGAGAGTCCAGAGCTTATCATCA 459  
 DB 2668 GCACCGCTTCTCTACCGCGGACAAATTAATTTGTGAGAGTGTCTTGTGCGCTCA 2509  
 QY 460 CTTTCTTTCTTTT 473  
 DB 2508 GGTTTTATCTT 2495  
 RESULT 15  
 AAA10594/c  
 ID AAA10594 standard; DNA: 10732 BP.  
 AC AAA10594;  
 XX 29-JUN-2000 (first entry)  
 DT Gene encoding a subunit of cellulose synthase.  
 DE Cellulose synthase: cellulose production; increase yield; ds.  
 KW *Vigna angularis*.  
 OS *Vigna angularis*.  
 XX JP2000060568-A.  
 XX 29-FEB-2000.  
 PD 26-AUG-1998; 98JP-0239998.  
 XX 26-AUG-1998; 98JP-0239998.  
 PR (MIZU/) MIZUNO K.  
 PA (CJIP) CJI PAPER CO.  
 XX WPI: 2000-342371/30.  
 DR P-PSDB; AAV85179.  
 XX  
 PT A gene encoding a cellulose synthetic equipment - for the improvement  
 PT in the amount of cellulose synthesised in a plant body

XX  
 PS Claim 2; Page 14-21; 32pp; Japanese.  
 XX  
 CC This sequence represents a gene encoding a subunit of the cellulose  
 CC synthase complex of *Vigna angularis*. The invention relates to subunits of  
 CC cellulose synthetic equipment, that can be used to increase the amount of  
 CC cellulose synthesised by a plant. The proteins and genes encoding them  
 CC can also be used to improve the properties of the cellulose being  
 CC produced by a plant.  
 XX  
 SO Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other:  
 Query Match 4.8%; Score 36.4; DB 21; Length 10732;  
 Best Local Similarity 18.3%; Pred. No. 3.9;  
 Matches 91; Conservative 167; Mismatches 237; Indels 3; Gaps 1;  
 QY 146 TTGCCAAGTGGCTCTTCCACAGGAAACACAAATATTTTGTCTTGTGACTGCTAG 205  
 DB 9510 BYSSDRCSRSDSTCNCYSCSDSRVSTYDACYTDAKTBCTYSDCNCSISRCYTNST 9451  
 QY 206 AACAAGGCGTGGGGGTGGGGAAGCAACTGATCGCCCTCTGAGAGACACCTGTGT 265  
 DB 9450 YGCSRTBSRNSRSTCCCTBTBSRNCCTDAYDANSTRIDYDACYSTYDASTBTBSICT 9391  
 QY 266 GCTGCGTGGCGCCAGGCTCTGCTGTGCTCTCTGAGCGGATGCTTGTGACTTGTCTA 325  
 DB 9390 BYSNSRYDAVSRYSCYTCYDYSSTCYTRCAKCTBCNSRRAKSTTTTTCYTCYT 9331  
 QY 326 CTTTTCACCTCTGAGCAGTCTCAGTCTCTCTGCTACCTTTTGTCTCCAGCTTCCCT 385  
 DB 9330 SRCYTTBCTCYTNCRSRGSRTSRSDAYSCNCSRSRSDYSTBCTYDACYCYYS 9271  
 QY 386 GCGGCTCGAATGACAGATACAGCAGACCTCTCTGTGACCGCTTGTGAGAGTCCAGAG 445  
 DB 9270 DCCYTCCNSTCTCTCCSRCTBCTTBCYRAYDYSDANSTRSRRAKSCCTTYAKTR 9211  
 QY 446 ACT--TTATCAATCCACTTTTCTTTTTCATTTGACCTGGGGGCCGAGGTTAAG 502  
 DB 9210 CSDSTCCTTNCYDANCCYCTSTNCSNCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 9151  
 QY 503 TACTTATCTGTCATCTCTGTGATGCAATGCGGAGTGGCGAGGCGCCCTTCTCCG 562  
 DB 9150 TCTYSRYSDACYSNCCCTBARCYTCSRDCSRVSTSRABTBTSTDCYTNCCNCCC 9091  
 QY 563 TCTTTCGCGCTCGAAGGGGTTCTCTTTATGCTCTTCCACCCACCCCTTCCCTCCCT 622  
 DB 9090 YTTBDCYTNCCAKSRGSRRTTCTGSDTTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 9031  
 QY 623 GCCCTCTGTTTGTGTC 640  
 DB 9030 ARYSTBCTTCTTSRSHRC 9013

Search completed: July 8, 2003, 03:11:10  
 Job time : 755.255 secs

PR 07-OCT-1999; 99US-0156029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159684.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 5.0%; Score 37.4; DB 21; Length 1522;  
Best Local Similarity 62.1%; Pred. No. 0.72;

Matches 59; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 281 TCTCCGTGCTGCTCTCTCTGAGCCGATGCTTGTGACTTGTCTTTTTCACCTAG 340  
DB 917 TCTCCTTTTACGCTTCCGCTTCGCTTCCTCTGACTTGATATTTTCTTCTCCG 858  
QY 341 CAGTCTCCAGTCTCTGCTGCTACTTTTGTCTCTCC 375  
DB 857 CCATCTCTCTTTTACAGCTACTTCTCTGCTCTCC 823

## RESULT 13

ABLO3779/c  
ID ABL03779 standard; cDNA: 4983 BP.

AC ABL03779;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 5819.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

XX WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PERE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR P-PSDB; ABB59676.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Claim 1; SEQ ID NO 5819; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA

CC sequences (AB101840-AB16175) and the encoded proteins

CC (ABB5737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 4983 BP; 1500 A; 1137 C; 1257 G; 1089 T; 0 other;

Query Match 4.9%; Score 37.2; DB 23; Length 4983;

Best Local Similarity 49.5%; Pred. No. 1.5;

Matches 96; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 280 GTCTCCTGTGCTGCTCTCTCTGAGCCGATGCTTGTGACTTGTCTTTTTCACCTGA 339  
DB 1196 GCCTTCTTTCCGCTTTTGTGTTTCCACACATATAGCTTGGCTTTTCTCGCTTA 1137  
QY 340 CCACTCTCCAGTCTCTCTGCTACTTTTGTGCTCCAGCTTCCCTGCGCTGAAATGC 399  
DB 1136 ACCGCTCCCAATTCCTCTGCGCAGAGATTAGCACTGAGCTTGGCTTGGCTTGA 1077  
QY 400 AGATACCGAGCTCCCTCTCTGAGCCGATGCTTGTGAGAGTCCAGAGACTTATCAATCCA 459  
DB 1076 GCAGCCGTGCTTACCGCGCACAATATATTGTGAGAGTGAATCTTCTTCCGCTCA 1017  
QY 460 CTTTCTTTTCTTTT 473  
DB 1016 GGTCTTTTAATCTT 1003

## RESULT 14

ABLO3778/c  
ID ABL03778 standard; cDNA: 7723 BP.

AC ABL03778;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 5816.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

XX WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PERE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139482.  
PR 18-JUN-1999; 99US-0139484.  
PR 18-JUN-1999; 99US-0139485.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140655.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148341.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 22-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151103.  
PR 31-AUG-1999; 99US-0151138.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.

```

PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144085.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144864.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145086.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148665.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.

```

```

PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160747.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

```

```

Query Match          5 0%; Score 37.4; DB 21; Length 1227;
Best Local Similarity 62.1%; Pred. No. 0.65;
Matches 59; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```

```

QY 281 TCTCTGTGTCGTTCTCTCTGAGCGATGCTTGACTTTCCTTTTCACGCTGAG 340
DB 919 TCTCTTTTCAGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTTCATATTTCTTCTCCG 860
QY 341 CAGTCTCCAGTTCCTCTGCTGCTACCTTTTGTCTCTCC 375
DB 859 CCATCTTCTTTTACAGCTACTCTCTGCTCTCC 825

```

## RESULT 12

AAC51060/c  
ID AAC51060 standard; DNA; 1522 BP.

XX AAC51060;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 67125.

XX Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

XX PN EPI03405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 9905-0121825.

PR 05-MAR-1999; 9905-0123180.

PR 09-MAR-1999; 9905-0123548.

P		XX		28-FEB-2000; 2000US-0515126.
R		PR		18-MAY-2000; 2000US-0577409.
X		XX		(HYSE-) HYSEQ INC.
P		PA		Tang YT, Liu C, Drmanac RT;
X		XX		WPI: 2001-514838/56.
D		DR		P-PDB: AAOI3574.
X		XX		Isolated nucleic acids and polypeptides, useful for preventing
P		PT		diagnosing and treating e.g. leukemia, inflammation and immune
X		XX		disorders -
SQ				Claim 1: SEQ ID NO 13565; 1399BP + Sequence Listing; English.
CC		XX		The invention relates to human polynucleotides (AA179941-AA193841) and
CC		CC		the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC		CC		cytokine, cell proliferation or cell differentiation or which may induce
CC		CC		production of other cytokines in other cell populations. The
CC		CC		polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC		CC		peptide therapy. The polypeptides have various cytokine-like activities,
CC		CC		e.g. stem cell growth factor activity, haematopoiesis regulating
CC		CC		activity, tissue growth factor activity, immunomodulatory activity and
CC		CC		activin/inhibin activity and may be useful in the diagnosis and/or
CC		CC		treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC		CC		inflammation.
CC	Note:			The sequence data for this patent did not form part of the printed
CC				spectification, but was obtained in electronic format directly from WIPO
CC				at ftp.wipo.int/pub/published_pcl_sequences.
SQ				Sequence 439 BP; 98 A; 91 C; 78 G; 172 T; 0 other:
Query Match	5.0%;	Score 38;	DB 22;	Length 439;
Best Local Similarity	53.3%;	Freq. No. 0.25;		
Matches	80;	Conservative	0;	Mismatches 70; Indels 0; Gaps 0;
OY	TTTTTTTCTTTTCATTTGGCCCCGCGAGGTTAAGTACTTATTCTGCATTGC	520		
Db				
	45 TTTTFTTTTTTTTTTTTTTTTTTGCGGGGGAGAGGGGAATACTGCCCCC	GSGGTG	104	
OY	521 TGTCGAATCACGAATGCCCTGAGGTGCACAGCCCTTTCCCTCTTTTGCGTCTGANAG	580		
Db				
	105 TTITTAACCCTTTTAAAAAAGGGGGAAAACCCCGGGGATTTTTAGGGGTGGGGGGG	164		
OY	581 GGTTCCTTTATGCTCTCACACC GCCCACCC	610		
Db				
	165 GGCCCCCTTTTTTTTTTTTCCGCCCAAAC	C	194	
RESULT 11				
AAC36552/C				
ID	AAC36552 standard; DNA; 1227 BP.			
XX	AAC36552;			
DT	17-OCT-2000 (first entry)			
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 14226.			
XX				
KM	Hybridisation assay; genetic mapping; gene expression control;			
XX	protein identification; signal transduction pathway;			
KW	metabolic pathway; promoter; termination sequence; ss.			
OS	Arabidopsis thaliana.			
PN	EP1033405-A2.			
PD	06-SEP-2000.			
PF	25-FEB-2000; 2000EP-0301439.			
XX				

FT		/number= "2"
FT	satellite	/1079..1156
FT		/*tag= e
FT	repeat_unit	/standard_name= "microsatellite"
FT		/1080..1083
FT		/*tag= f
FT	satellite	/1565..1599
FT		/*tag= g
FT	repeat_unit	/standard_name= "microsatellite"
FT		/1566..1567
FT	exon	/*tag= h
FT		/1625..1919
FT		/*tag= i
FT	intron	/number= "3"
FT		/1920..2511
FT		/*tag= j
FT		/number= "3"
FT	exon	/2512..2829
FT		/*tag= k
FT	intron	/number= "4"
FT		/2830..4471
FT		/*tag= l
FT		/number= "4"
FT	satellite	/3088..3138
FT		/*tag= m
FT	repeat_unit	/standard_name= "microsatellite"
FT		/3088..3091
FT		/*tag= n
FT	satellite	/3779..3780
FT		/*tag= p
FT	repeat_unit	/3874..3901
FT		/*tag= q
FT	satellite	/3797..3823
FT		/*tag= o
FT	repeat_unit	/standard_name= "microsatellite"
FT		/3779..3780
FT		/*tag= r
FT	repeat_unit	/4017..4272
FT		/*tag= s
FT	satellite	/standard_name= "microsatellite"
FT		/4017..4022
FT	repeat_unit	/*tag= t
FT		/4472..4645
FT	exon	/*tag= u
FT		/number= "5"
FT	intron	/4646..5355
FT		/*tag= v
FT		/number= "5"
FT	satellite	/4914..4982
FT		/*tag= w
FT	repeat_unit	/standard_name= "microsatellite"
FT		/4915..4916
FT		/*tag= x
FT	exon	/5356..5484
FT		/*tag= y
FT	intron	/number= "6"
FT		/5485..6697
FT		/*tag= z
FT	exon	/number= "6"
FT		/6898..7066
FT		/*tag= aa
FT	intron	/number= "7"
FT		/7067..7404
FT		/*tag= ab
FT	exon	/number= "7"
FT		/7405..8300
FT		/*tag= ac
XX		/number= "8"
XX	NO200125442-A1.	
PD	12-APR-2001.	

[illegible]

PR 28-JUL-2000; 2000US-221607P.  
 PR 02-MAY-2001; 2001US-287724P.  
 XX  
 XX (COMP-) COMPUGEN INC.  
 PA Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;  
 PI WPI; 2002-257383/30.  
 XX  
 XX  
 PT New oligonucleotide libraries comprising oligonucleotides which  
 PT selectively hybridize to mRNAs transcribed from a transcription unit of  
 PT a genome, useful for detecting tissue-, pathology-, and  
 PT developmental-specific genes  
 PS Example 1; SEQ ID 31650; 47pp; English.  
 XX  
 CC The present invention describes oligonucleotide libraries for detecting  
 CC messenger RNAs that populate a (sub-)transcriptome, where the  
 CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple  
 CC transcription units that populate a genome. The library comprises  
 CC several oligonucleotides, each capable of hybridizing selectively to a  
 CC set of messenger RNAs transcribed from a given transcription unit of  
 CC the genome, which encodes one or more messenger RNA splice variants.  
 CC The oligonucleotide libraries are useful for detecting mRNAs from a  
 CC biological sample, in expression profiling studies, in qualitatively or  
 CC quantitatively characterizing the corresponding transcriptome, and in  
 CC detecting RNA transcripts and splice variants of human or animal  
 CC transcripts. The libraries may also be used as specialized mini  
 CC libraries to detect transcripts of a sub-transcriptome under a  
 CC particular biological or pathological state, and so allowing the  
 CC detection of tissue- and pathology-specific genes such as those genes  
 CC only expressed in specific tissue under a specific pathological  
 CC condition; to detect developmental specific genes; and to detect RNA  
 CC transcripts and splice variants of a transcriptome of a patient suffering  
 CC from a particular disorder. ABN27253 to ABN59589 represent  
 CC oligonucleotide sequences from rats, humans and mice, which are used in  
 CC the exemplification of the present invention.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 CC  
 SQ Sequence 60 BP; 14 A; 8 C; 28 G; 10 T; 0 other;  
 Query Match 8.0%; Score 60; DB 24; Length 60;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-08;  
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 72 ACGTGGAGTTAGCGAGTTCGAGACTCGATGTCGAGCCAGAAAGGTCGAG 131  
 Db 1 ACGTGGAGTTAGCGAGTTCGAGACTCGATGTCGAGCCAGAAAGGTCGAG 60  
 RESULT 8  
 ABL33323/c  
 ID ABL33323 standard; DNA; 9741 BP.  
 XX  
 AC ABL33323;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 1296.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianemic; cytosinatic; nootropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX

PN W0200200928-A2.  
 XX  
 XX 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-EP07537.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-130909/17.  
 XX  
 PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation  
 PS Claim 1; SEQ ID NO 1296; 32pp + Sequence Listing; German.  
 XX  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 CC  
 SQ Sequence 9741 BP; 2623 A; 195 C; 1863 G; 5060 T; 0 other;  
 Query Match 5.2%; Score 39.4; DB 24; Length 9741;  
 Best Local Similarity 59.3%; Pred. No. 0.44;  
 Matches 67; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
 QY 640 CCCAANAACAGTCTCTAAGCTTTCGATGTCGAGCGGAAATTAACCTGACC 699  
 Db 662 CCTAANAACAGTCTCTAAGCTTTCGATGTCGAGCGGAAATTAACCTGACC 603  
 QY 700 CGTTGCAAAATGAACTTTTTCCTGATCCTGACACCTGTTTAACT 752  
 Db 602 TTTATAAATACGAAATTTTCACTTTTTCCTGATCCTGACACCTGTTTAACT 550  
 RESULT 9  
 AAH25711  
 ID AAH25711 standard; DNA; 8300 BP.  
 XX  
 AC AAH25711;  
 XX  
 DT 14-AUG-2001 (first entry)  
 XX  
 DE Murine DNA polymerase lambda gene.  
 XX  
 KW Mouse; human; DNA polymerase lambda; DNA repair; isoform; cancer;  
 KW neurodegenerative disease; gene therapy; pol kappa; POLT; chromosome 19;  
 KW aging; immunosuppression; psoriasis; arthritis; graft rejection; ds.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT 1..58  
 FT /tag= a  
 FT /number= "1"  
 FT intron  
 FT 59..827  
 FT /tag= b  
 FT /number= "1"  
 FT exon  
 FT 828..964  
 FT /tag= c  
 FT /number= "2"  
 FT intron  
 FT 965..1624  
 FT /tag= d



CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
CC DNA chip analysis as given in the specification, and comparing  
CC the expression level to an expression level in an unactivated

[illegible]

XX 20-JUL-2001, 2001WO-IB01903.  
PF  
XX

ID AAT61405 standard; cDNA; 2887 BP.  
 XX AAT61405:  
 XX  
 XX 29-OCT-1997 (first entry)  
 DE MACH isoform alpha coding sequence.  
 XX  
 KM MACH: MORT-1 binding protein; mediator of receptor toxicity; cell death;  
 KM antibody; FAS ligand receptor; FAS-R; death domain region; septic shock;  
 KM tumour necrosis factor; tumour; HIV-infection; oligodendrocyte death;  
 KM apoptosis/programmed cell death; p55-R; graft rejection; acute hepatitis;  
 KM autoimmune disease; multiple sclerosis; AIDS-inhibited T-cell suicide;  
 KM TNF; therapy: ss.  
 XX  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 291..1731  
 FT /tag= a  
 FT /product= MORT-1 binding protein (MACH) isoform alpha  
 XX  
 PN MO9703998-A1.  
 PD 06-FEB-1997.  
 PD 14-JUN-1996; 96MO-US10521.  
 PF 16-APR-1996; 96IL-0117932.  
 PR 16-JUL-1995; 95IL-0114615.  
 PR 17-AUG-1995; 95IL-0114986.  
 PR 14-SEP-1995; 95IL-0115319.  
 PR 27-SEP-1995; 95IL-0116588.  
 XX  
 PA (WEIN/) WEINMURZEL H.  
 PA (YEDA ) YEDA RES & DEV CO LTD.  
 PI Boldin M, Goltssev YV, Goncharov T, Wallach D:  
 DR MPI: 1997-132570/12.  
 DR P-PSDB; AAM11891.  
 XX  
 PT New DNA encoding MACH protein that interacts with MORT-1 protein -  
 PT to mediate intracellular effects of FAS or TNF receptors, partic.  
 PT for regulating apoptosis in tumours, virus-infected cells etc.  
 XX  
 PS Example 3; Page 116-117; 163pp; English.  
 XX  
 CC AAT61396, and AAT61405-T61411 represent coding sequences for different  
 CC isoforms of MACH. MACH is a binding protein for the mediator of receptor  
 CC toxicity (MORT-1) protein. MORT-1 binds to the FAS ligand receptor  
 CC (FAS-R) death domain region, and triggers part of the cell death  
 CC signalling cascade in mammalian cells. Vectors containing these  
 CC sequences, the encoded proteins, and antibodies (Ab) against them are  
 CC used to modulate the effect of FAS-R ligand or TNF on cells that carry  
 CC FAS-R or p55-R. This is specifically for treating tumours, HIV-infected  
 CC cells or other diseased cells, by control of apoptosis/programmed cell  
 CC death. The encoded proteins are mediators of the cell death pathway  
 CC initiated by TNF and FAS-R binding, i.e. it mimics or enhances the  
 CC effect of MORT-1 where increased cytotoxicity is required. To inhibit the  
 CC effect of MORT-1, e.g. in cases of septic shock, graft rejection and  
 CC acute hepatitis, sequences encoding antisense molecules or ribozymes, or  
 CC Ab against the protein, are used. Compounds that inhibit MACH are  
 CC potentially useful for controlling MACH activity e.g. in cases of  
 CC autoimmune disease, oligodendrocyte death in multiple sclerosis or  
 CC AIDS-inhibited T-cell suicide. The MACH proteins can also be used to  
 CC isolate and characterise other proteins and receptors involved in  
 CC signalling and for Ab production. The Ab can be used to purify the new  
 CC proteins and for diagnosis of conditions involving abnormal function of  
 CC FAS-R mediated cellular effects.  
 XX  
 SQ Sequence 2887 BP; 853 A; 595 C; 625 G; 814 T; 0 other;

Query Match 16.4%; Score 123.6; DB 18; Length 2887;  
 Best Local Similarity 93.5%; Pred. No. 2,8e-27;  
 Matches 129; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
 QY 13 TATTGAAGCTAAAGAACTTCTTCTGAGACCTTTCCACCCCTTCCCTGAGCA 72  
 DB 144 TATTGAAGCTAAAGAACTTCTTCTGAGACCTTTCCACCCCTTCCCTGAGCA 203  
 QY 73 CGTGAGTTAGGCACTGAGGAGCTGAGAGCTGATGTCGACAGAAAGGTTGAGC 132  
 DB 204 CATTGAGTTAGGCACTGAGGAGCTGAGAGCTGATGTCGACAGAAAGGTTGAGC 263  
 QY 133 GGGTGAAGTCCCTGTTGCC 150  
 DB 264 GGATTATATTCCTGCTGCC 281  
 RESULT 5  
 AAD03915  
 ID AAD03915 standard; DNA; 2887 BP.  
 XX  
 AC AAD03915;  
 XX  
 DT 02-JUL-2001 (first entry)  
 XX  
 DE Human caspase-8 DNA.  
 XX  
 KM Human: caspase-8; MACH; FLICE; Mch3;  
 KM cysteine protease; apoptosis; caspase expression cassette; metastasis;  
 KM tumour; cathepsin B; urokinase; proliferation; gene therapy; ds.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 292..1731  
 FT /tag= a  
 FT /product= "Human caspase-8"  
 XX  
 PN WO200129232-A2.  
 PD 26-APR-2001.  
 PD 19-OCT-2000; 2000MO-US28941.  
 PR 20-OCT-1999; 99US-0160559.  
 PR 14-AUG-2000; 2000US-0225564.  
 XX  
 PA (SCIO-) SCIOS INC.  
 PI Cordell B, Li Y;  
 DR MPI: 2001-290920/30.  
 DR P-PSDB; AAE00605.  
 XX  
 PT Novel fusion polypeptide comprising first and second caspase subunit  
 PT separated by cleavage site not associated in nature with caspase  
 PT subunit, useful for cloning gene encoding enzymes involved in  
 PT proteolytic cleavage  
 XX  
 PS Disclosure: Fig 15; 116pp; English.  
 XX  
 CC The present sequence is a DNA encoding human Caspase-8 also known as  
 CC MACH, FLICE and Mch3. Caspases are a family of cysteine proteases, that  
 CC participate in the initiation and execution of apoptosis. Caspases exist  
 CC as pro-enzymes, activated by cleavage into a large and small subunit,  
 CC occurring after specific aspartic acid residues within the pro-enzyme  
 CC sequence. The present invention relates to a method for functional  
 CC cloning of genes encoding proteins or enzymes involved in proteolytic  
 CC cleavage. The invention is based on the use of caspase expression  
 CC cassettes comprising the coding sequence of a proteolytic cleavage site  
 CC flanked by sequences encoding two caspase subunits. A fusion polypeptide  
 CC comprising a first and a second caspase subunit, separated by a cleavage  
 CC site not associated in nature, is useful for cloning gene encoding

PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Barash SC, Ruben SM;  
 XX  
 XX WPI; 2001-483426/52.  
 XX  
 XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -  
 XX  
 PS Disclosure; SEQ ID NO 33029; 3071bp + Sequence Listing; English.  
 XX  
 CC AA65951 to AA654702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AA652170 to AA651921. (I) have cytosolic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased

CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC cancers and cancer metastases of hematopoietic-derived cells. AA654703  
 CC to AA65794 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AA654942 to AA654950 and AA652169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 45017 BP; 12919 A; 10037 C; 9566 G; 12495 T; 0 other:  
 Query Match 70.4%; Score 529.8; DB 22; Length 45017;  
 Best Local Similarity 91.0%; Pred. No. 9, se-152;  
 Matches 675; Conservative 0; Mismatches 42; Indels 20; Gaps 10;  
 13 TATTGAAGTAAAGAAAGAACTTCTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 72  
 25499 TATTGAAGTAAAGAAAGAACTTCTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 25440  
 QY 73 CGTGAAGTAAAGAAAGAACTTCTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 132  
 Db 25439 CGTGAAGTAAAGAAAGAACTTCTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 25380  
 QY 133 GGGTGAAGTAAAGAAAGAACTTCTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 192  
 Db 25379 GGGTGAAGTAAAGAAAGAACTTCTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 25320  
 QY 193 TGACTTGTCTAGAAAGAAAGAACTTCTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 252  
 Db 25319 TGACTTGTCTAGAAAGAAAGAACTTCTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 25260  
 QY 253 GGAACACTCT-GGTGTGCTGCTGAGAGCCTTCCGAGAGCCTTCCGAGAGCCTTCCGAGAGC 311  
 Db 25259 GGAACACTCTGAGAGCCTTCCGAGAGCCTTCCGAGAGCCTTCCGAGAGCCTTCCGAGAGC 25200  
 QY 312 CTTTGAAGTAAAGAAAGAACTTCTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 371  
 Db 25199 CTTTGAAGTAAAGAAAGAACTTCTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 25140  
 QY 372 CTGCAAGTAAAGAAAGAACTTCTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 431  
 Db 25139 CTGCAAGTAAAGAAAGAACTTCTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 25080  
 QY 432 GGAAGTAAAGAAAGAACTTCTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 491  
 Db 25079 GGAAGTAAAGAAAGAACTTCTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 25021  
 QY 492 CGAGGTTAAGTAAAGAAAGAACTTCTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 551  
 Db 25020 CGAGGTTAAGTAAAGAAAGAACTTCTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 24962  
 QY 552 CCCCTTCCCTGCTGAGAGCCTTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 611  
 Db 24961 CCCCTTCCCTGCTGAGAGCCTTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 24905  
 QY 612 TTCCCTTCCCTGCTGAGAGCCTTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 671  
 Db 24904 TTCCCTTCCCTGCTGAGAGCCTTCCGAGAGCCTTCCACCCCTCCCTGCTGAGCA 24852  
 QY 672 GTGAGTTGCGGAAATTAACCTGACAGCCTTTCGAAATGAAGCTTTTGTGATCC 731  
 Db 24851 GTGAGTTGCGG--AATTATCTGTATC--TTTGCAAAATGTA--TTTCTTTGTGANG 24799  
 QY 732 TGTACTGCTGTTTGA 748  
 Db 24798 TGTACTGCTGTTTGA 24782  
 RESULT 4  
 AAT61405

Db 24300 GGGTGAAGTGGCTGTGTCACAGGAGGCGCTCTTCCACAGAAACACAAATATTTTGTCT 24241  
QY 193 TGACTTGCTCTAGAAACAGGCGCTGTGGGGGTGGGAGCAACTTGGATGTGCCCTTTCGA 252  
Db 24240 TGACTTGCTCTAGAAACAGGCGCTGTGGGGGTGGGAGCAACTTGGATGTGCCCTTTCGA 24181  
QY 253 GGACACCTCT-GGTGTGCTGGCCAGCGCTCTGTGGTTCCTCTGTAGCCGAGTGC 311  
Db 24180 GGACACCTCTGGTGTGCTGGCCAGCGCTCTGTGGTTCCTCTGTAGCCGAGTGC 24121  
QY 312 CTTTGACTTGTCTACTTTTTCACCTGTAGCAGTCTCCAGTTCCTCTGTACTTTTGTGC 371  
Db 24120 CTCTGACTTGTCTACTTTTTCACCTGTAGCAGTCTCCAGTTCCTCTGTACTTTTGTGC 24061  
QY 372 CTCACAGCTCCCTGGCGGCTGTGAATGACATACAGGAGTCCCTGTGTGAGCCGTTT 431  
Db 24060 CTCACAGCTCCCTGGCGGCTGTGAATGACATACAGGAGTCCCTGTGTGAGCCGTTT 24001  
QY 432 GGAGAGTCCAGAGACTTTATCAATCCACTTTTTCCTTTTTCATTTTGGCCCTGGGGGC 491  
Db 24000 GGAGAGTCCAGAGACTTTATCAATCCACTTTTTCCTTTTTCATTTTGGCCCT 23942  
QY 492 CGACGTTAAGTCTTATTTCTCTCATTTCTGTGCAATCAGATGCCCTGAGGTGACAG 551  
Db 23941 CGACGTTAAGTCTTATTTCTCTCATTTCTGTGCAATCAG-ATGCCCTGAGGTGACAG 23883  
QY 552 CCCCCTTCCCTCTTCCGCGCTCTGAAGGAGTTCCTTTATGTCCTCCACCCACCT 611  
Db 23882 CCCCCTTCCCTCTTCCGCGCTCTGAAGGAGTTCCTTTATGTCCTCCACCCACCT 23826  
QY 612 TTCCCTTCCCTCTTCCGCGCTCTGTGTTGTGCCCCAAAAAACAAGTCTCTAAAGTTTGCAT 671  
Db 23825 TGCCCTTCCCTCTTCCGCGCTCTGTGTTGTGCCCCAAAAAACAAGTCTCTAA- TGTTTGCAT 23773  
QY 672 GTGATTCGCGGGAATTAATCAACCGACCCGTTGCAAAATGAATTTTTCATGTC 731  
Db 23772 GTGATTCGCGGGAATTAATCAACCGACCCGTTGCAAAATGA- TTTTTCATGTC 23720  
QY 732 TGTAACACTGTTTTTTA 748  
Db 23719 TGTAACACTGTTTTTTA 23703

RESULT 3  
AAK77217/c  
ID AAK77217 standard; DNA: 45017 BP.  
XX  
AC AAK77217;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32029.  
XX  
KW Human: immune; haematopoietic; immune/haematopoietic antigen; cancer:  
XX cytosolic; gene therapy; vaccine; metastasis; ds.  
OS Homo sapiens.  
XX  
FN W0200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01354.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0216889.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218299.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225279.  
PR 14-AUG-2000; 2000US-0225281.  
PR 22-AUG-2000; 2000US-0225286.  
PR 22-AUG-2000; 2000US-0225287.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226281.  
PR 22-AUG-2000; 2000US-0226286.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231966.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 12-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239835.  
PR 13-OCT-2000; 2000US-0239837.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.

```

PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249255.
PR 17-NOV-2000; 2000US-0249257.
PR 17-NOV-2000; 2000US-0249259.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM,
PI WPI: 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX Disclosure; SEQ ID NO 32028; 3071pp + Sequence Listing; English.
PS
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK67694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
XX Sequence 43938 BP; 12772 A; 9656 C; 9206 G; 12304 T; 0 other;
SQ
Query Match 70.4%; Score 529.8; DB 22; Length 43938;
Best Local Similarity 91.6%; Pred. 9.2e-152;
Matches 675; Conservative 0; Mismatches 42; Indels 20; Gaps 10;

OY 13 TATTGAAGTAAGAAAGAACTTCTCTCTGGAGCCTTCCACCCCTCCCTGCTGAGCA 72
DB 24420 TATTGAAGTAAGAAAGAACTTCTCTCTGGAGCCTTCCACCCCTCCCTGCTGAGCA 24361
OY 73 CGTGAGCTTAGGCGAGCTTAGGCGAGCTCGGAGACCTGCATGGTGCAGAAAGGTGAGC 132
DB 24360 CGTGAGCTTAGGCGAGCTTAGGCGAGCTCGGAGACCTGCATGGTGCAGAAAGGTGAGC 24301
OY 133 GGGTAGTGCTCTTGCCAAAGGTGAGCTCTTCAACAGAAACCAATATTGTTGTTCT 192

```

Pr comprising the CASP8 gene that results in inactivation of the gene  
XX  
PS Claim 21, Page 97-98; 107pp; English.  
XX  
CC Inactivation of a caspase-8 (CASP8) gene is determined by detecting a  
CC modification of CASP8 genomic DNA that results in inactivation of the  
CC gene. CASP8, a cysteine protease, is part of the death inducing signaling  
CC complex (DISC) associated with the Fas receptor. CASP8 is inactivated in  
CC cancers, and plays a role of a tumour suppressor gene. The CASP8 promoter  
CC region sequences, in particular Region 1 and Region 2, are crucial to the  
CC design and execution of the genomic methylation PCR analysis of CASP8  
CC gene inactivation. Methylation PCR can be used to examine even minute  
CC amounts of patient material to demonstrate whether the CASP8 gene  
CC expresses an mRNA and protein product. The promoter Region 1 sequence is  
CC located upstream (5') to exon 1, which is the alternatively-spliced 5'  
CC untranslated region (UTR) that is less commonly used. The promoter  
CC Region 2 sequence is located downstream (3') of exon 1 and upstream of  
CC exon 2, which is the more commonly used 5' UTR. The CASP8 gene has been  
CC localized to human chromosome 2q33-34. The methods are used to diagnose  
CC or prognose cancer. Cancer is treated by administering a vector that  
CC expresses a gene encoding functional CASP8 in cells. The cancer that  
CC is diagnosed or treated is a tumour in which a MYC gene is amplified,  
CC such as a neuroblastoma. Aggressive neuroblastoma, juvenile neuroblastoma  
CC (preferred), small-cell lung carcinoma, non-small-cell lung carcinoma,  
CC colorectal carcinoma, or uterine cervical carcinoma can be diagnosed with  
CC the new method. A kit for screening for a compound that induces  
CC death-receptor-mediated apoptosis in cells containing an inactivated  
CC CASP8 gene is also provided.  
XX  
XX  
SQ Sequence 753 BP; 140 A; 204 C; 171 G; 238 T; 0 other;  
  
Query Match 100.0%; Score 753; DB 21; Length 753;  
Best Local Similarity 100.0%; Pred. No. 2.3e-221;  
Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 AATTAGACCGCGTATGAAGTAAGAAAGTCTCTCTGGGAGCCTTCCACCCCTT 60  
DB 1 AATTAGACCGCGTATGAAGTAAGAAAGTCTCTCTGGGAGCCTTCCACCCCTT 60  
  
OY 61 CCCTGCTGAGACAGTGGAGTTAGGAGGAGTGGAGACTGCGATGATGTCACAG 120  
DB 61 CCCTGCTGAGACAGTGGAGTTAGGAGGAGTGGAGACTGCGATGATGTCACAG 120  
  
OY 121 AAGGSGTGAGGCGGGTGAGGCTGTGGCAAGTGCGCTTTCACAGAAACCAAT 180  
DB 121 AAGGSGTGAGGCGGGTGAGGCTGTGGCAAGTGCGCTTTCACAGAAACCAAT 180  
  
OY 181 AATTGTTGTTCTGACTGTCTAGAAACAGGCTGTGGGAGTGGAGCAACTTGAT 240  
DB 181 AATTGTTGTTCTGACTGTCTAGAAACAGGCTGTGGGAGTGGAGCAACTTGAT 240  
  
OY 241 CTGCGCTTCTGAGAGACCTCTGCTGCTGGCCAGAGTCTCTGTGTGTTCTCTC 300  
DB 241 CTGCGCTTCTGAGAGACCTCTGCTGCTGGCCAGAGTCTCTGTGTGTTCTCTC 300  
  
OY 301 TGAGCCGAGGACCTTACTTGTCTTCTTTCACAGAGTCTCCAGTCTCTGCT 360  
DB 301 TGAGCCGAGGACCTTACTTGTCTTCTTTCACAGAGTCTCCAGTCTCTGCT 360  
  
OY 361 ACCTTTTGTCTCCAGACTCCCTCGCGCTCGAGATGACAGATACAGACCTCTTG 420  
DB 361 ACCTTTTGTCTCCAGACTCCCTCGCGCTCGAGATGACAGATACAGACCTCTTG 420  
  
OY 421 TGAGACCGGTTGAGAGTCCAGAGACTTATCAATCCACTTTTCTTTTCAATTG 480  
DB 421 TGAGACCGGTTGAGAGTCCAGAGACTTATCAATCCACTTTTCTTTTCAATTG 480  
  
OY 481 GCCCTGGGGCCGACGATTAACTTTATCTGTCACTCTGCAATCAGCAATGCCCT 540  
DB 481 GCCCTGGGGCCGACGATTAACTTTATCTGTCACTCTGCAATCAGCAATGCCCT 540  
  
OY 541 GAGGTGACAGCCCTTCCCTCTTCCGCTCTCAAGGGGTTTCTTTATGTTTCC 600  
DB 541 GAGGTGACAGCCCTTCCCTCTTCCGCTCTCAAGGGGTTTCTTTATGTTTCC 600

DB 541 GAGGTGACAGCCCTTCCCTCTTCCGCTCTCAAGGGGTTTCTTTATGTTTCC 600  
OY 601 ACCCCACACCTTCCCTCCCTGCGCCCTGTGTTTGTGGCCAAATAACAGTCTCTTA 660  
DB 601 ACCCCACACCTTCCCTCCCTGCGCCCTGTGTTTGTGGCCAAATAACAGTCTCTTA 660  
OY 661 ACGTTTGGATGTGATTCGGGAAATTAACCTGCACCCGTTTGCAAAATGAACTTTT 720  
DB 661 ACGTTTGGATGTGATTCGGGAAATTAACCTGCACCCGTTTGCAAAATGAACTTTT 720  
OY 721 TTTTGTGATCCTGTACACTGTTTTTTACCTT 753  
DB 721 TTTTGTGATCCTGTACACTGTTTTTTAACTT 753  
  
RESULT 2  
AAK77216/C  
ID AAK77216 standard; DNA; 43938 BP.  
XX  
XX AAK77216;  
AC  
XX  
XX 07-NOV-2001 (first entry)  
DT  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32028.  
XX  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200157182-A2.  
PN  
XX  
PD 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01354.  
PF  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225457.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226881.  
PR 22-AUG-2000; 2000US-0226886.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:03:20 ; Search time 751.255 Seconds  
(without alignments)  
2257.229 Million cell updates/sec

Title: US-09-477-082-2

Perfect score: 753  
Sequence: 1 aatagacgcgtatgaaa.....tacactgttttttaacctt 753

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SID82/gcgcdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SID82/gcgcdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SID82/gcgcdata/geneseq/geneseq-emb1/NA2001.DAT:\*  
23: /SID82/gcgcdata/geneseq/geneseq-emb1/NA2002.DAT:\*  
24: /SID82/gcgcdata/geneseq/geneseq-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	753	100.0	753	21	AA51793
2	529.8	70.4	43938	22	AAK77216
3	529.8	70.4	45017	22	AAK77217
4	123.6	16.4	2887	18	AA61405
5	123.6	16.4	2887	18	AA61405
6	123.6	16.4	2887	18	AA61405
7	60	8.0	60	24	ABK84723
8	39.4	5.2	9741	24	ABN58902
9	38.4	5.1	8300	22	ABL33323
					AAH25711

10	38	5.0	439	22	AA193505	Human polynucleoti
11	37.4	5.0	1227	21	AA36552	Arabidopsis thalia
12	37.4	5.0	1522	21	AA351060	Arabidopsis thalia
13	37.2	4.9	4983	23	AB103779	Drosophila melanog
14	37.2	4.9	7723	23	AB103778	Drosophila melanog
15	36.4	4.8	10732	21	AA10594	Gene encoding a su
16	35.8	4.8	700	22	AAH92818	Human inflammatory
17	35.8	4.8	700	22	AAH92819	Human inflammatory
18	35.8	4.8	9743	22	AA104517	Human reproductive
19	35.8	4.8	90220	24	ABR83576	Human CDNA differe
20	35.4	4.7	2291	23	AB110669	Drosophila melanog
21	35.4	4.7	3079	23	AB108820	Drosophila melanog
22	35.4	4.7	4453	23	AB110668	Drosophila melanog
23	35.2	4.6	2603	24	ABR65157	Arabidopsis CDNA e
24	35	4.6	6146	22	ABA07752	Human ovarian and
25	35	4.6	6146	22	AA103541	Human reproductive
26	34.8	4.6	744	22	AA195563	Human neuroblastom
27	34.8	4.6	6741	21	AA110585	Gene encoding a su
28	34.8	4.6	30310	24	AA138253	Genomic DNA from t
29	34.6	4.6	6277	21	AA176859	Human OREF ORF2414
30	34.4	4.6	448	22	AA180552	Human polynucleoti
31	34.4	4.6	2779	24	ABR84714	Human CDNA differe
32	34.4	4.6	3059	22	AA158335	Human polynucleoti
33	34.4	4.6	3068	22	AA158336	Human polynucleoti
34	34.4	4.6	3071	22	AA158337	Human polynucleoti
35	34.4	4.6	3134	22	AA158334	Human polynucleoti
36	34.4	4.6	3713	21	AA178160	Human cancer assoc
37	34.4	4.6	126512	24	ABN83429	Human transporter
38	34.2	4.5	1494	24	ABR63596	Rat sequence diffe
39	34	4.5	143068	21	AAE21105	Human low adenosin
40	34	4.5	143068	21	AAE21105	Human low adenosin
41	34	4.5	143068	21	AAE21105	Human low adenosin
42	34	4.5	143068	21	AAE21105	Human low adenosin
43	34	4.5	143068	24	AA65150	Human adenosine re
44	34	4.5	143068	24	AB168124	Ovary cancer relat
45	34	4.5	12740	21	AA35151	Human adenosine re
					AAE21273	Human low adenosin

## ALIGNMENTS

AA51793	standard; DNA; 753 BP.
AA51793	31-OCT-2000 (first entry)
AA51793	CASP8 promoter Region 2.
AA51793	CASP8; caspase-8; promoter; Region 2; inactivation; methylation;
AA51793	Cysteine protease; tumour suppressor; chromosome 2q33-34; neuroblastoma;
AA51793	cancer; death receptor; apoptosis; cytosolic; gene therapy; ss.
AA51793	Homo sapiens.
AA51793	WO2000039347-A1.
AA51793	06-JUL-2000.
AA51793	30-DEC-1999; 99WO-0531280.
AA51793	31-DEC-1998; 98US-0114308.
AA51793	(SUJ-) ST JUDE CHILDREN'S RES HOSPITAL.
AA51793	Kidd VJ, Lahti JM, Teitz T;
AA51793	WPI: 2000-452423/39.
AA51793	Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or
AA51793	prognosing cancer, comprises detecting a modification of genomic DNA









TITLE  
 JOURNAL  
 REFERENCE  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Tellrod, B., Thomas, N., Thomas, S.,  
 Umanil, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, X., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.  
 Direct Submission  
 2 (bases 1 to 128941)  
 Morley, K.C.  
 Direct Submission  
 Submitted (11-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 128941)  
 Worley, K.C.  
 Direct Submission  
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 12, 2002 this sequence version replaced gi:18847091.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GRX  
 Center clone name: CH230-257E5  
 ----- Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye: 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 84896 bases at least Q40  
 Consensus quality: 89146 bases at least Q30  
 Consensus quality: 92809 bases at least Q20

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 55 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1422: contig of 1422 bp in length  
 \* 1423 1522: gap of unknown length  
 \* 1523 2531: contig of 1009 bp in length  
 \* 2532 2631: gap of unknown length  
 \* 2632 3933: contig of 1302 bp in length  
 \* 3934 4033: gap of unknown length  
 \* 4034 5147: contig of 1114 bp in length  
 \* 5148 5247: gap of unknown length  
 \* 5248 6375: contig of 1128 bp in length  
 \* 6376 6475: gap of unknown length  
 \* 6476 7682: contig of 1207 bp in length  
 \* 7683 7782: gap of unknown length  
 \* 7783 8958: contig of 1176 bp in length  
 \* 8959 9058: gap of unknown length  
 \* 9059 10124: contig of 1066 bp in length  
 \* 10125 10224: gap of unknown length  
 \* 10225 11431: contig of 1207 bp in length  
 \* 11432 11531: gap of unknown length  
 \* 11532 12623: contig of 1092 bp in length  
 \* 12624 12723: gap of unknown length  
 \* 12724 14127: contig of 1404 bp in length  
 \* 14128 14227: gap of unknown length  
 \* 14228 15363: contig of 1136 bp in length  
 \* 15364 15463: gap of unknown length  
 \* 15464 16626: contig of 1163 bp in length

\* 16627 16726: gap of unknown length  
 \* 16727 17904: contig of 1178 bp in length  
 \* 17905 18004: gap of unknown length  
 \* 18005 19198: contig of 1194 bp in length  
 \* 19199 19298: gap of unknown length  
 \* 19299 20550: contig of 1152 bp in length  
 \* 20551 20550: gap of unknown length  
 \* 20551 21588: contig of 1038 bp in length  
 \* 21589 21689: gap of unknown length  
 \* 21689 22697: contig of 1009 bp in length  
 \* 22698 22797: gap of unknown length  
 \* 22798 23963: contig of 1166 bp in length  
 \* 23964 24063: gap of unknown length  
 \* 24064 26414: contig of 2351 bp in length  
 \* 26415 26514: gap of unknown length  
 \* 26515 27791: contig of 1277 bp in length  
 \* 27792 27891: gap of unknown length  
 \* 27892 28601: contig of 1710 bp in length  
 \* 28602 29701: gap of unknown length  
 \* 29702 31394: contig of 1653 bp in length  
 \* 31395 31494: gap of unknown length  
 \* 31495 32917: contig of 1423 bp in length  
 \* 32918 33017: gap of unknown length  
 \* 33018 34819: contig of 1802 bp in length  
 \* 34820 34919: gap of unknown length  
 \* 34920 37158: contig of 2239 bp in length  
 \* 37159 37258: gap of unknown length  
 \* 37259 39046: contig of 1788 bp in length  
 \* 39047 39146: gap of unknown length  
 \* 39147 41095: contig of 1949 bp in length  
 \* 41096 41196: gap of unknown length  
 \* 41197 43465: contig of 2270 bp in length  
 \* 43466 43565: gap of unknown length  
 \* 43566 44580: contig of 1015 bp in length  
 \* 44581 44680: gap of unknown length  
 \* 44681 47162: contig of 2482 bp in length  
 \* 47163 47262: gap of unknown length  
 \* 47263 48007: contig of 1645 bp in length  
 \* 48008 49007: gap of unknown length  
 \* 49009 50894: contig of 1887 bp in length  
 \* 50895 50994: gap of unknown length  
 \* 50995 52332: contig of 1938 bp in length  
 \* 52333 53032: gap of unknown length  
 \* 53033 54663: contig of 1631 bp in length  
 \* 54664 54763: gap of unknown length  
 \* 54764 56965: contig of 2202 bp in length  
 \* 56966 57065: gap of unknown length  
 \* 57066 59764: contig of 2699 bp in length  
 \* 59765 59864: gap of unknown length  
 \* 59865 62414: contig of 2550 bp in length  
 \* 62415 62514: gap of unknown length  
 \* 62515 65241: contig of 2727 bp in length  
 \* 65242 65341: gap of unknown length  
 \* 65342 67570: contig of 2229 bp in length  
 \* 67571 67670: gap of unknown length  
 \* 67671 70493: contig of 2823 bp in length  
 \* 70494 70593: gap of unknown length  
 \* 70594 72869: contig of 2376 bp in length  
 \* 72870 73069: gap of unknown length  
 \* 73070 75496: contig of 2427 bp in length  
 \* 75497 77828: gap of unknown length  
 \* 77829 77928: contig of 2232 bp in length  
 \* 77929 81696: gap of unknown length  
 \* 81697 81796: contig of 3768 bp in length  
 \* 81797 84238: contig of 2442 bp in length  
 \* 84239 84338: gap of unknown length  
 \* 84339 86916: contig of 2578 bp in length  
 \* 86917 87016: gap of unknown length  
 \* 87017 90937: contig of 3921 bp in length  
 \* 90938 91037: gap of unknown length  
 \* 91038 95061: contig of 4024 bp in length  
 \* 95062 95161: gap of unknown length

CDS .916

Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,

KEYWORDS cell death; MACH-alpha-1; protease.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 2887)  
 AUTHORS Boldin, M.P., Goncharov, T.M., Goltsev, Y.V. and Wallach, D.  
 TITLE Involvement of MACH, a novel MORT1/FADD-interacting protease, in Fas/APO-1- and TNF receptor-induced cell death  
 JOURNAL Cell 85 (6), 803-815 (1996)  
 MEDLINE 96279826  
 PUBMED 8681376  
 REFERENCE 2 (bases 1 to 2887)  
 AUTHORS Wallach, D.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-MAY-1996) D. Wallach, The Weizmann Institute, Dept of Membrane Research & Biophysics, Rehovot 76100, ISRAEL  
 REFERENCE 3 (bases 1 to 2887)  
 AUTHORS Wallach, D.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-JUN-1996) D. Wallach, The Weizmann Institute, Dept of Membrane Research & Biophysics, Rehovot 76100, ISRAEL  
 REFERENCE 4 (bases 1 to 2887)  
 AUTHORS Manduzato, S., Brasseur, F., Andry, G., Boon, T. and van der Bruggen, P.  
 TITLE A CASP-8 mutation recognized by cytolytic T lymphocytes on a human head and neck carcinoma  
 JOURNAL J. Exp. Med. 186 (5), 785-793 (1997)  
 MEDLINE 97419196  
 PUBMED 9271594  
 FEATURES  
 source location/Qualifiers  
 1..2887  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /tissue\_type="thymus"  
 292..1731  
 /function="protease"  
 /note="triggers cell death"  
 /codon\_start=1  
 /product="MACH-alpha-1"  
 /protein\_id="CA66853.1"  
 /db\_xref="GI:2440071"  
 /db\_xref="SWISS-PROT:O14790"  
 /translation="MDESRNLYDIGEQLSDSLASLKLFLSLDYIPQRKQEPKDALML  
 FORLEKRLLESNLSFLKELFRINRLDLITLTKREMERELQTPGSAQISAYR  
 VMLYQISEEVSSELSFKFLQEEISCKLDDNNLDIFLEMERVITIGSKDIL  
 KRYCAQINKSLKLIINDYEFSEKSSLESGDPDESNDEELGVTITSDSPEDSE  
 SOLIDKIVOMKSPRGVCLINNNPARKAREVPKLSIRDRNGHLDGALTTPPE  
 LHEIKPDDCTVEOYIEIKIYOLMDHNMDCFTCCILSHDGIITGTDGCAPIY  
 ELTSGFTGKCSIAKPKVFFIACOGNPKGTPVETDSEQPIRENDLSSPOTRY  
 IPRADFLGKATVNVNVCYVRNPASTWTIQSLQSLRERCRGDILITLTVNIEV  
 SNDDKKNMGKMPQPTFTLRKRLVPPSP"

BASE COUNT 853 a 595 c 625 g 814 t  
 ORIGIN

Query Match 16.4%; Score 123.6; DB 9; Length 2887;  
 Best local Similarity 93.5%; Pred. No. 4.3e-27;  
 Matches 123; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 13 TATTGAAGTAAAGAACTTCTCTCTGGAGAGCCCTTCCACCCCTCTCCTGAGACA 72  
 Db 144 TATTGAAGTAAAGAACTTCTCTCTGGAGAGCCCTTCCACCCCTCTCCTGAGACA 203

QY 73 CGTGGAGTTAGGAGGAGTGGAGACTCGAGACTGCGATGGTCCAGGAAGGTTGAGAC 132  
 Db 204 CGTGGAGTTAGGAGGAGTGGAGACTCGAGACTGCGATGGTCCAGGAAGGTTGAGAC 263

QY 133 GGGTGGAGTGGTGGTGGC 150  
 Db 264 GGATTAATATCTCTCTGCC 281

RESULT 9  
 AF422929  
 LOCUS 777 bp mRNA linear PRI 03-APR-2002  
 DEFINITION Homo sapiens clone 4 procaspase-8 (CASP8) mRNA, partial cds;  
 accession alternatively spliced.  
 ACCESSION AF422929  
 VERSION AF422929.1 GI:19401529  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 777)  
 AUTHORS Breckenridge, D.G., Nguyen, M., Kuppig, S., Reith, M. and Shore, G.C.  
 TITLE The procaspase-8 isoform, procaspase-8L, recruited to the BAP31 complex at the endoplasmic reticulum  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4331-4336 (2002)  
 MEDLINE 21327603  
 PUBMED 11917123  
 REFERENCE 2 (bases 1 to 777)  
 AUTHORS Breckenridge, D.G., Nguyen, M., Kuppig, S., Reith, M. and Shore, G.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-SEP-2001) Biochemistry, McGill University, 3655 Promenade Sir William Osler, Montreal, PQ H3G1Y6, Canada  
 FEATURES  
 source location/Qualifiers  
 1..777  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="4"  
 1..>777  
 /gene="CASP8"  
 /note="caspase precursor; alternatively spliced"  
 /codon\_start=1  
 /product="procaspase-8"  
 /protein\_id="AA87632.1"  
 /db\_xref="GI:19401530"  
 /translation="MDESRNLYDIGEQLSDSLASLKLFLSLDYIPQRKQEPKDALML  
 FORLEKRLLESNLSFLKELFRINRLDLITLTKREMERELQTPGSAQISAYR  
 VMLYQISEEVSSELSFKFLQEEISCKLDDNNLDIFLEMERVITIGSKDIL  
 KRYCAQINKSL"

BASE COUNT 218 a 156 c 209 g 194 t  
 ORIGIN

Query Match 16.2%; Score 121.8; DB 9; Length 777;  
 Best local Similarity 98.4%; Pred. No. 1.4e-26;  
 Matches 123; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 13 TATTGAAGTAAAGAACTTCTCTCTGGAGAGCCCTTCCACCCCTCTCCTGAGACA 72  
 Db 144 TATTGAAGTAAAGAACTTCTCTCTGGAGAGCCCTTCCACCCCTCTCCTGAGACA 203

QY 73 CGTGGAGTTAGGAGGAGTGGAGACTCGAGACTGCGATGGTCCAGGAAGGTTGAGAC 132  
 Db 204 CGTGGAGTTAGGAGGAGTGGAGACTCGAGACTGCGATGGTCCAGGAAGGTTGAGAC 263

QY 133 GGGTGG 137  
 Db 264 GGATG 268

RESULT 10  
 I66494  
 LOCUS 7218 bp DNA linear PAT 28-DEC-1997  
 DEFINITION Sequence 14 from patent US 5670367.  
 ACCESSION I66494  
 VERSION I66494.1 GI:2724471  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 7218)



OY		73	CGTGGAGTTTGAGGCACTGGAGAACTGCAGATGGTGGCAGGAAGGGGTGAGC	133
OY		1682	CGTGAGTTTAGCGAGGTTTAGGGGACTCGGAGACACTGCCAATGTTGCCAGAAAGGTTGAGC	1741
OY		133	GGGTGAAGTCCTGTGGCCAGGTGGCCTCTTCAACAGGAACAACAATAATTTTGTTC	192
Db		1742	GGGTGAAGTGGCTGTGGCCAGAGTGGCCTCTTCAACAGGAACAACAATAATTTTGTTC	1801
OY		193	TGATTTCCTCAGAAACAGGGGCTGGGGGGTGGGAAGAACAATTGATCTCCCTTCGA	252
Db		1802	TGACTTTCCTCAGAAACAGGGGCTGGGGGGTGGGAAGAACAATTGATCTCCCTTCGA	1861
OY		253	GGACACCTCT-GGTGCTGECCTGGCCAGATCCATCTGTGTGGTTTTCTCTAGCCGATG	311
Db		1862	GGACACCTCTGGGTGGCTGGCCAGGAGTCCATCTGTGTGGTTTTCTCTAGCCGATG	1921
OY		312	CTTTGACCTTGGCTACTTTTTCACCTCAGACAGTCCACAGTCCCTCTGCTACTTTTGTG	371
Db		1922	CTCTGACTTGGCTACTTTTTCACCTCAGACAGTCCACAGTCCCTCTGCTACTTTTGTG	1981
OY		372	CTCCAAGCTTCCCTGGCCGCTCGAATCGAATACACGAGCTCCCTCTGTGAGCCGTT	431
Db		1982	CTCCAAGCTTCCCTGGCCGCTCGAATCGAATACACGAGTCCCTCTGTGAGCCGTT	2041
OY		432	GGAGAGTCCGAAACACTTAATCATCCACTTTTTTCTTTTATTTTGGCCCTGGGGC	491
Db		2042	GGAGAGTCCGAAACACTTAATCATCCACTTTTTTCTTTTATTTTGGCCCTGGGGC	2100
OY		492	CGAGGTTTAAGTACTTATCTGTGATTTGTGCAATCACAGAACGCCGTGAGTGCACAG	551
Db		2101	CGAGGTTTAAGTACTTATTTAT-----CTGTGCAATCAG-ANGCCTGAGGTGCACAG	2151
OY		552	CCCCCTTCCCCTCTTTTGGCGTCTCGAAGAGGGTTTCCTTTATGTCTTCCACCOCCT	611
Db		2152	CCCCCTTCCCCTCTTCCGTGTC--TGAAGGGTTTCTCTTTATCTC-TCCACCCCACT	2208
OY		612	TTCCTTCCTCCCTGCCCCCTGTGTTTGTGTGGCCCAAACCAATTCCTCFAACTTTGCAT	671
Db		2209	TGCCCTCCCTGCCCCCTCTCTGTGTT-----CCCCAAGACATTTCTCTA--TGTTTTCAT	2261
OY		672	GTCGATTCGGGAAAATTAACCTGCACCCGTTTGCAAAATGACATTTTTTTGTATCC	731
Db		2262	GTCGATTCGGG--AATTTATCTGTATC--TTTGCAAAATGTA--TTTTCCTTTTGTATG	2314
OY		732	TGTACTGCTGTTTTTA	748
Db		2315	TGTACTGCTTTTTTA	2331
RESULT 4				
AF210257			982 bp DNA linear	PRI 06-JUN-2000
LOCUS			Homo sapiens caspase 8 gene, 5' flanking region.	
DEFINITION			AF210257	
ACCESSION			AF210257.1	GI:8272474
VERSION				
KEYWORDS				
SOURCE			Homo sapiens.	
ORGANISM			Homo sapiens	
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi;	
			Mammalia; Eutheraia; Primates; Carnivora; Cetartihidi; Homnidae; Homo.	
REFERENCE				
AUTHORS			Teitz,T., Mel,T., Valentine,M.B., Vanin,E.F., Grenet,J.,	
			Valentine,V.A., Beim,F.G., Look,A.T., Lahli,J.M. and Kidd,V.J.	
JOURNAL			Nat. Med. 6 (5), 529-535 (2000)	
MEDLINE			20264561	
PUBMED			10802708	
REFERENCE			2 (bases 1 to 982)	
AUTHORS			Teitz,T., Mel,T., Valentine,M.B., Vanin,E.F., Grenet,J.,	
			Valentine,V.A., Beim,F.G., Look,A.T., Lahli,J.M. and Kidd,V.J.	
TITLE			Direct Submission	

JOURNAL		Submitted (01-DEC-1999) Tumor Cell Biology, St. Jude Children's Research Hospital, 332 N. Lauderdale, Memphis, TN 38105, USA	
FEATURES		Location/Qualifiers	
SOURCE		1..982	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/chromosome="2"	
		/map="2q33-q34"	
		1..982	
misc_feature		/note="5', flanking region of caspase 8 gene"	
BASE COUNT		198 a 235 c 241 g 308 t	
ORIGIN			
Query Match	64.9%; Score 488.4; DB 9; Length 982;		
Best Local Similarity	98.5%; Pred. No. 2,1e-141;		
Matches 514;	Conservative 0; Mismatches 6; Indels 2; Gaps 2;		
QY	13 TATTGAAAGTAAAGAAACCTTCTTCTGCGAGCCCTTCCACCCCTTCCCTGCTGAGCA 72		
Db	459 TATTGAAAGTGAAGAAACCTTCTTCTGCGAGCCCTTCCACCCCTTCCCTGCTGAGCA 518		
QY	73 CGTGAGGTTGGCGAGGTTAGGGGAGCTCGGAAGCTCGCATGGTGGCAGAAAGGTTGAGC 132		
Db	519 CGTGAGGTTAGCGAGGTTAGGGGAGCTCGGAAGCTCGCATGGTGGCAGAAAGGTTGAGC 578		
QY	133 GGGTGAAGTGCCTGTGTGCCAAGGTGGCCCTCTTCAACAGAGAAACCAATATTATTGTTTCT 192		
Db	579 GGGTGAAGTGCCTGTGTGCCAAGGTGGCCCTCTTCAACAGAGAAACCAATATTATTGTTTCT 638		
QY	193 TGACTTTGCTCTAGAAACAGGGCTGTGGGGGTTGGGGAAAGCACTTGATCTGCCCTTTCGA 252		
Db	639 TGACTTTGCTCTAGAAACAGGGCTGTGGGGGTTGGGGAAAGCACTTGATCTGCCCTTTCGA 698		
QY	253 GGACACCCCTCTGGAGCTGCGTGGCCAGAGGTCCTGATGGTGTCTCTGAGCCGATGC 311		
Db	699 GGACACCCCTCTGGAGCTGCGTGGCCAGAGGTCCTGATGGTGTCTCTGAGCCGATGC 758		
QY	312 CTTTGACTTTGCTACTTTTTCACCTCTGAGCAGTGTCCAGTTCCTGTGTACTTTTGTGTC 371		
Db	759 CTTTGACTTTGCTACTTTTTCACCTCTGAGCAGTGTCCAGTTCCTGTGTACTTTTGTGTC 818		
QY	372 CTCCAAGCTTCCCGCCGCGCCCTCGAATGAGATATACGAGACCTCTTCTGTGAGCCGTTT 431		
Db	819 CTCCAAGCTTCCCGCCGCGCCCTCGAATGAGATATACGAGACCTCTCTCTGTGAGCCGTTT 878		
QY	432 CGAGAGTCCAGAAAGACTTTATCAATCCACTTTTTCCTTTTCATTGTGGCCTCGGGGC 491		
Db	879 CGAGAGTCCAGAAAGACTTTATCAATCCACTTTTTCCTTTTCATTGTGGCCTCGGGGC 937		
QY	492 CGACGGTTAACTATTATCTGTGATCTTCTGTGCAATCAGCA 533		
Db	938 CGACGGTTAACTATTATCTGTGATCTTCTGTGCAATCAGCA 979		
RESULT 5	AF422925	2503 bp	mrna
LOCUS	AF422925		linear
DEFINITION	Homo sapiens procaspase-8L (CASP8) mRNA, complete cds;		PRI 03-APR-2002
ACCESSION	AF422925		
VERSION	AF422925.1	GI:19401518	
KEYWORDS			
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 2503)		
AUTHORS	Breckenridge, D.G., Nguyen, M., Kuppig, S., Beth, M. and Shore, G.C.		
TITLE	The procaspase-8 isoform, procaspase-8L, recruited to the BAP31		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4331-4336 (2002)		
MEDLINE	21927603		
PUBMED	11917123		

```

misc-feature      5194..5761
                  /note="similar to Homo sapiens EST BF529656
                  (NID:g1617019)"
misc-feature      5237..6019
                  /note="similar to Homo sapiens EST AU130508
                  (NID:g10990862)"
misc-feature      5238..6043
                  /note="similar to Mus musculus EST BF166846
                  (NID:g1047198)"
misc-feature      5326..5653
                  /note="similar to Homo sapiens EST BG390376
                  (NID:g13283624)"
misc-feature      5361..6051
                  /note="similar to Homo sapiens EST AU135897
                  (NID:g10996436)"
misc-feature      5361..5812
                  /note="similar to EST BG662638 (NID:g14213176)"
misc-feature      5420..6048
                  /note="similar to Homo sapiens EST BE788113
                  (NID:g10209311)"
misc-feature      5551..6327
                  /note="similar to Homo sapiens EST AU130771
                  (NID:g10991125)"
misc-feature      5615..55984
                  /note="similar to Homo sapiens EST AA280659 (NID:g1923454)

Query Match      70.4%; Score 529.8; DB 9; Length 181150;
Best Local Similarity 91.6%; Pred. 3.9e-154;
Matches 675; Conservative 0; Mismatches 42; Indels 20; Gaps 10;

QY      13 TATTGAAGTAAAGAACTTCTCTCTGGAGCCTTCCACCCCTTCCCTGCTGAGCA 72
DB      159209 TATTGAAGTAAAGAACTTCTCTCTGGAGCCTTCCACCCCTTCCCTGCTGAGCA 159150

QY      73 CGTGGAGTAGGAGAGTGGGGAGCTCGAGAGCTGGATGATGCGCAGAGAGGTGGAGC 132
DB      159149 CGTGGAGTAGGAGAGTGGGGAGCTCGAGAGCTGGATGATGCGCAGAGAGGTGGAGC 159090

QY      133 GGGTAGTGCCTGTTGCCAAGTGCCCTTTCAACAGAAACCAATATTTTGTCT 192
DB      159089 GGGTAGTGCCTGTTGCCAAGTGCCCTTTCAACAGAAACCAATATTTTGTCT 159030

QY      193 TGACTGCTCTAGAAACAGGGGCTGGGGGGGAGAGCACTGGATGCTGCCCTTCA 252
DB      159029 TGACTGCTCTAGAAACAGGGGCTGGGGGGGAGAGCACTGGATGCTGCCCTTCA 158970

QY      253 GGACACCTCT-GGTGTGCTGGGCCAGGTCCTGTGTGTTCTCTCTGAGCGATGC 311
DB      158969 GGACACCTCTGGGTGCTGGGCCAGGTCCTGTGTGTTCTCTCTGAGCGATGC 158910

QY      312 CTTTGACTTGTCTACTTTTCACTTGACAGCTCTCCACTTCTCTGCTACTCTTTTGTG 371
DB      158909 CTTTGACTTGTCTACTTTTCACTTGACAGCTCTCCACTTCTCTGCTACTCTTTTGTG 158850

QY      372 CTCACAGCTTCCCTGCGCCCTGCAATGATACAGAGCTCCCTCTGTGACCCGTT 431
DB      158849 CTCACAGCTTCCCTGCGCCCTGCAATGATACAGAGCTCCCTCTGTGACCCGTT 158790

QY      432 GGAAGATCCAGAAAGACTTATCAATCCACTTTTCTTTTCAATTGAGCCCTGGGGC 491
DB      158789 GGAAGATCCAGAAAGACTTATCAATCCACTTTTCTTTTCAATTGAGCCCT-GGGGC 158731

QY      492 CGAGGTTAGTACTTATTTCTGTCTGTCGATGCAATGAGCCCTGAGGTGACAGC 551
DB      158730 CGAGGTTAGTACTTATTTCTGTCTGTCGATGCAATGAGCCCTGAGGTGACAGC 158672

QY      552 CCCCTTCCCTCTTTCGGGTCCTGAAGGGGTTTCCTTTATGTCCTCCACCCCAACCT 611
DB      158671 CCCCTTCCCTCTTTCGGGTCCTGAAGGGGTTTCCTTTATGTC-TCACACCCCAACCT 158615

QY      612 TTCCCTCCCTGCTCTGTTTGTGTCCTCAAAACAGTTCTTAACGTTTTCAT 671
DB      158614 TTCCCTCCCTGCTCTGTTTGTGTCCTCAAAACAGTTCTTAAC-TCGTTTTCAT 158562

```

```

QY      672 GTGATTCGCGAAATTAACCTGCACCCGTTTCGAAATGAACTTTTGTGATCC 731
DB      158561 GTGATTCGTC---AATTACTGTATC--TTGCAAAATGTA--TTTCTCTTGTATG 158509

QY      732 TGTACACTGCTTTTGA 748
DB      158508 TGTACACTGCTTTTGA 158492

RESULT 3
AB03898052
LOCUS      AB03898052 6502 bp DNA linear PRI 23-FEB-2001
DEFINITION Homo sapiens gene for caspase-8, exon 3, exon 4.
ACCESSION AB038981
VERSION    AB038981.1 GI:12862688
KEYWORDS
SEGMENT
SOURCE
ORGANISM   Homo sapiens DNA.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS    1 Hadano, S., Yanagisawa, Y., Skaug, J., Pichter, K., Nasir, J.,
            Martindale, D., Koop, B.E., Scherer, S.W., Nicholson, D.W.,
            Rouleau, G.A., Ikeda, J.-E. and Hayden, M.R.
            Cloning and characterization of three novel genes, ALS2CR1,
            ALS2CR2, and ALS2CR3, in the juvenile amyotrophic lateral sclerosis
            (ALS2) critical region at chromosome 2q33-q34: candidate genes for
            ALS2
            Genomics 71 (2), 200-213 (2001)
            21100893
MEDLINE
REFERENCE
AUTHORS    Hadano, S., Ikeda, J. and Hayden, M.R.
TITLE      Direct Submission
JOURNAL    Submitted (24-FEB-2000) Shinji Hadano, Tokai University, The
            Institute of Medical Sciences, Bohseidai, Isehara, Kanagawa
            259-1193, Japan (E-mail:shinji.engi.med.u-tokai.ac.jp,
            Tel:+81-463-91-5095, Fax:+81-463-91-4993)
FEATURES
source
            1..6502
            /organism="Homo Sapiens"
            /db_xref="taxon:9606"
            /chromosome="2"
            /map="2q33"
            <1..1478
            /note="intron is approximately 24.0 kbp long"
            /number=2
            1479..1743
            /product="caspase-8"
            /note="CDS is reported in Acc#AB038985
            alternative 5'UTR"
            /number=3
            1744..3924
            /number=3
            3925..3970
            /product="caspase-8"
            /note="CDS is reported in Acc#AB038985
            alternative 5'UTR"
            /number=4
            3971..>6502
            /note="intron is approximately 5.9 kbp long"
            /number=4
            1718 a 1361 c 1504 g 1919 t
BASE COUNT
ORIGIN
Query Match      66.5%; Score 500.6; DB 9; Length 6502;
Best Local Similarity 90.2%; Pred. No. 3.8e-145;
Matches 665; Conservative 0; Mismatches 44; Indels 28; Gaps 11;

QY      13 TATTGAAGTAAAGAACTTCTCTCTGGAGCCTTCCACCCCTTCCCTGCTGAGCA 72
DB      1622 TATTGAAGTAAAGAACTTCTCTCTGGAGCCTTCCACCCCTTCCCTGCTGAGCA 1681

```







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:07:09 ; Search time 2674.33 Seconds  
(without alignments)  
8194.363 Million cell updates/sec

Title: US-09-477-082-2  
Perfect score: 753  
Sequence: 1 aatagacgcgcgtatgaaa.....tacacgtgttttaacctt 753

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

GenEmbl: \*  
1: gb\_da: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_da: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htg\_mus: \*  
34: em\_htg\_pln: \*  
35: em\_htg\_rtd: \*  
36: em\_htg\_man: \*  
37: em\_htg\_vtl: \*  
38: em\_sy: \*  
39: em\_htgo\_hum: \*  
40: em\_htgo\_mus: \*  
41: em\_htgo\_other: \*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	529.8	70.4	179216	2 AC074016	AC074016 Homo sapi
2	529.8	70.4	181150	9 AC007256	AC007256 Homo sapi
3	500.6	66.5	6502	9 AB038980s2	AB038981 Homo sapi
4	488.4	64.9	982	9 AF210257	AF210257 Homo sapi
5	123.6	16.4	2503	9 AF422925	AF422925 Homo sapi
6	123.6	16.4	2887	6 AR211527	AR211527 Sequence
7	123.6	16.4	2887	6 AR344419	AR344419 Sequence
8	123.6	16.4	2887	9 HSBACB61	X98172 H.sapiens m
9	121.8	16.2	777	9 AF422929	AF422929 Homo sapi
10	83.8	11.1	7218	6 AF6494	AF6494 Sequence 14
11	62.2	8.3	1723	9 AF380342	AF380342 Homo sapi
12	45	6.0	128941	2 AC110348	AC110348 Rattus no
13	45	6.0	146383	2 AC116367	AC116367 Oryza sat
14	44.6	5.9	61246	2 AL359032_4	Continuation (5 of
15	44.6	5.9	145763	2 AC064809	AC064809 Homo sapi
16	44.6	5.9	163677	9 AL390844	AL390844 Human DNA
17	44	5.8	237727	2 AC099291	AC099291 Rattus no
18	43.4	5.8	168975	2 AC117170	AC117170 Rattus no
19	43	5.7	124156	9 AL596132	AL596132 Human DNA
20	43	5.7	183617	2 AC007524	AC007524 Homo sapi
21	42.8	5.7	128941	2 AC110348	AC110348 Rattus no
22	42.4	5.6	177407	2 AC091553	AC091553 Trypanoso
23	41.4	5.5	143394	2 AC103397	AC103397 Rattus no
24	41.4	5.5	171749	2 AC112440	AC112440 Rattus no
25	41.4	5.5	182360	2 AC107184	AC107184 Rattus no
26	41.4	5.5	183151	2 AC112825	AC112825 Rattus no
27	41.2	5.5	125020	9 AF429315	AF429315 Homo sapi
28	41	5.4	44429	2 AC083762	AC083762 Homo sapi
29	41	5.4	56342	2 AC103270	AC103270 Rattus no
30	41	5.4	230290	2 AC101859	AC101859 Mus muscu
31	41	5.4	303091	2 AC084799	AC084799 Mus muscu
32	41	5.4	176691	2 AC097904	AC097904 Rattus no
33	40.6	5.3	106582	2 AC114068	AC114068 Rattus no
34	40.2	5.3	110000	2 AC084053_1	Continuation (2 of
35	40.2	5.3	182539	2 AC110338	AC110338 Rattus no
36	40.2	5.3	157665	10 AC069018	AC069018 Mus muscu
37	40	5.3	161396	2 AC095069	AC095069 Rattus no
38	40	5.3	173418	2 AC110354	AC110354 Rattus no
39	39.8	5.3	969	11 CNS066DL	AL401455 T3 end of
40	39.8	5.3	128144	2 AC113915	AC113915 Rattus no
41	39.8	5.3	142333	2 AC107270	AC107270 Rattus no
42	39.8	5.3	167470	2 AP005545	AP005545 Oryza sat
43	39.8	5.3	171457	2 AC095247	AC095247 Rattus no
44	39.8	5.3	196912	10 AL513346	AL513346 Mouse DNA
45	39.6	5.3	196912	10 AL513346	AL513346 Mouse DNA

## ALIGNMENTS

RESULT 1  
AC074016 179216 bp DNA linear HTG 09-AUG-2001  
LOCUS AC074016  
DEFINITION Homo sapiens chromosome 2 clone RP11-155P18, WORKING DRAFT  
ACCESSION AC074016  
VERSION AC074016.5 GI:15144340  
KEYWORDS HTG; HTGS PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 179216)  
AUTHORS Waterston,R.H.  
TITLE The sequence of Homo sapiens clone



PRIOR FILLING DATE: 2001-04-06  
2000-04-06

FEATURE:  
NAME/KEY: unsure  
LOCATION: (15565, 15570, 15574, 15580, 15585, 15586, 15589, 15590, 17255)  
US-10-239-676-53

Query Match  
Best Local Similarity 4.9%; Score 32.6; DB 9; Length 17421;  
Best Local Similarity 60.9%; Pred. No. 20;  
Matches 53; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 582 TTACGCTGGAGGAGTGTTCACAGGTTCTCCCTTTATCTTTGTTGTTTTTTT 641  
DB 13184 TTTAGAGAGAGAGAGAGTGTTCATTAATGTTGTTTGGAGTTTATTTTTTTT 13243  
QY 642 CGAGCATGGGGGTAAATAAGCGCT 668  
DB 13244 TTGGGTTGGAGGTTAAATATGAGCT 13270

RESULT 8  
US-08-755-235-3/c  
Sequence 3, Application US/08755235  
Publication No. US20030059423A1  
GENERAL INFORMATION:  
APPLICANT: Stern, David M.  
APPLICANT: Schmidt, Ann Marie  
TITLE OF INVENTION: METHOD FOR TREATING SYMPTOMS OF DIABETES  
FILE REFERENCE: 0575/50159  
CURRENT APPLICATION NUMBER: US/08/755,235  
CURRENT FILING DATE: 1996-11-22  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
LENGTH: 1405  
TYPE: DNA  
ORGANISM: Human  
US-08-755-235-3

Query Match  
Best Local Similarity 4.8%; Score 32.4; DB 7; Length 1405;  
Best Local Similarity 64.9%; Pred. No. 5.8;  
Matches 48; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 176 AGCCACAGGGGCTACTTACCCAGTCCGGGAGGAGAGAGGGGCTGTCTGACTT 235  
DB 839 AGGGGAGGGGACACATCTTCATCCAGTGTGAGGAGAGGGCTGGCAGGAGCTT 780  
QY 236 CAGTGTGAGGTTT 249  
DB 779 CACAGTCAAGGTT 766

RESULT 9  
US-08-905-709-3/c  
Sequence 3, Application US/08905709  
Patent No. US20010039256A1  
GENERAL INFORMATION:  
APPLICANT: Stern, David  
APPLICANT: Schmidt, Ann M.  
TITLE OF INVENTION: A METHOD TO PREVENT ACCELERATED  
TITLE OF INVENTION: ATHEROSCLEROSIS USING (SRAGE) SOLUBLE RECEPTOR FOR  
TITLE OF INVENTION: ADVANCED GLYCATION ENDPRODUCTS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/905,709  
FILING DATE: 05-AUG-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0575/52876  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0526  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1405 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-905-709-3

Query Match  
Best Local Similarity 4.8%; Score 32.4; DB 8; Length 1405;  
Best Local Similarity 48.6%; Pred. No. 5.8;  
Matches 48; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 176 AGCCACAGGGGCTACTTACCCAGTCCGGGAGGAGAGAGGGGCTGTCTGACTT 235  
DB 839 AGGGGAGGGGACACATCTTCATCCAGTGTGAGGAGAGGGCTGGCAGGAGCTT 780  
QY 236 CAGTGTGAGGTTT 249  
DB 779 CACAGTCAAGGTT 766

RESULT 10  
US-09-822-830A-512  
Sequence 512, Application US/09822830A  
Patent No. US20020142952A1  
GENERAL INFORMATION:  
APPLICANT: Genetics Institute, Inc.  
APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Fechtel, Kim  
APPLICANT: Agostino, Michael J.  
APPLICANT: Howes, Steven H.  
APPLICANT: Resnick, Richard J.  
APPLICANT: Gulukota, Kamalakari  
APPLICANT: Graham, James R.  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
FILE REFERENCE: GIN 6402  
CURRENT APPLICATION NUMBER: US/09/822,830A  
CURRENT FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/195,604  
PRIOR FILING DATE: 2000-04-06  
NUMBER OF SEQ ID NOS: 631  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 512  
LENGTH: 1426  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-822-830A-512

Query Match  
Best Local Similarity 4.8%; Score 32.2; DB 10; Length 1426;  
Best Local Similarity 48.6%; Pred. No. 6.8;  
Matches 88; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 481 TGTTCAGTACATCTCTGTTCGCTTTAGAGTAAGTTTACCTGCAGTCTCTCT 540  
DB 903 TGGCAGATGATTTATCTTACCTGCTGATGATGATGATGATGATGATGATGAT 962  
QY 541 GTGGTAGATTTTCTCTCTCTCGAGACAGATTCGCTTTACGCTGAGGAGAGTG 600  
DB 963 GTTGTATTATTTACATTTGTCACGTTTGTAGAGAAATGTTAACAATGTTAAACTC 1022

;; TITLE OF INVENTION: 97 Human secreted proteins  
;; FILE REFERENCE: P2028p1  
;; CURRENT APPLICATION NUMBER: US/09/892,877  
;; CURRENT FILING DATE: 2001-06-28  
;; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658  
;; PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10  
;; NUMBER OF SEQ ID NOS: 461  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 103  
;; LENGTH: 704  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: SITE  
;; LOCATION: (287)  
;; OTHER INFORMATION: n equals a,t,g, or c  
US-09-892-877-103

Query Match 4.9%; Score 32.6; DB 9; Length 704;  
Best Local Similarity 52.6%; Pred. No. 3.4;  
Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 149 TTTTCTGACGATCTACCCGATTTGACGACAGGGCTGACTTACCCAGTCGGCGGG 208  
DB 525 TATACTGAGGCGACTATAGTAGTAAATCTTTATTTATTTTGGAGAC 584  
QY 209 AGGAGGAGAGGCGTGTGCTGACTGAGTGTGAGTTGATCAAGGCAAGGGAAC 268  
DB 585 AGGAGGCGCTCTGCTTTGTTTCAGACTGGAGTGCAGTGTGTGATCATGTGCTCATTTGCAAC 644  
QY 269 TTCCATTTCCAGAC 283  
DB 645 TTGAACCTCTGGGC 659

RESULT 6  
US-09-948-783-103  
;; Sequence 103, Application US/09948783  
;; Publication No. US20030100051A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Ruben et. al.  
;; TITLE OF INVENTION: 97 Human secreted proteins  
;; FILE REFERENCE: P2028p2  
;; CURRENT APPLICATION NUMBER: US/09/948,783  
;; CURRENT FILING DATE: 2001-09-10  
;; PRIOR APPLICATION NUMBER: 60/231,846  
;; PRIOR FILING DATE: 2000-09-11  
;; PRIOR APPLICATION NUMBER: 09/892,877  
;; PRIOR FILING DATE: 2001-06-28  
;; PRIOR APPLICATION NUMBER: 09/437,658  
;; PRIOR FILING DATE: 1999-11-10  
;; PRIOR APPLICATION NUMBER: PCT/US99/09847  
;; PRIOR FILING DATE: 1999-05-06  
;; PRIOR APPLICATION NUMBER: 60/085,093  
;; PRIOR FILING DATE: 1998-05-12  
;; PRIOR APPLICATION NUMBER: 60/085,094  
;; PRIOR FILING DATE: 1998-05-12  
;; PRIOR APPLICATION NUMBER: 60/085,105  
;; PRIOR FILING DATE: 1998-05-12  
;; PRIOR APPLICATION NUMBER: 60/085,180  
;; PRIOR FILING DATE: 1998-05-12  
;; PRIOR APPLICATION NUMBER: 60/085,927  
;; PRIOR FILING DATE: 1998-05-18  
;; PRIOR APPLICATION NUMBER: 60/085,906  
;; PRIOR FILING DATE: 1998-05-18  
;; PRIOR APPLICATION NUMBER: 60/085,924  
;; PRIOR FILING DATE: 1998-05-18  
;; PRIOR APPLICATION NUMBER: 60/085,922  
;; PRIOR FILING DATE: 1998-05-18  
;; PRIOR APPLICATION NUMBER: 60/085,921  
;; PRIOR FILING DATE: 1998-05-18  
;; PRIOR APPLICATION NUMBER: 60/085,923  
;; PRIOR FILING DATE: 1998-05-18

;; PRIOR APPLICATION NUMBER: 60/085,925  
;; PRIOR FILING DATE: 1998-05-18  
;; PRIOR APPLICATION NUMBER: 60/085,928  
;; PRIOR FILING DATE: 1998-05-18  
;; PRIOR APPLICATION NUMBER: 60/085,920  
;; PRIOR FILING DATE: 1998-05-18  
;; NUMBER OF SEQ ID NOS: 465  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 103  
;; LENGTH: 704  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: SITE  
;; LOCATION: (287)  
;; OTHER INFORMATION: n equals a,t,g, or c  
US-09-948-783-103

Query Match 4.9%; Score 32.6; DB 9; Length 704;  
Best Local Similarity 52.6%; Pred. No. 3.4;  
Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 149 TTTTCTGACGATCTACCCGATTTGACGACAGGGCTGACTTACCCAGTCGGCGGG 208  
DB 525 TATACTGAGGCGACTATAGTAGTAAATCTTTATTTATTTTGGAGAC 584  
QY 209 AGGAGGAGAGGCGTGTGCTGACTGAGTGTGAGTTGATCAAGGCAAGGGAAC 268  
DB 585 AGGAGGCGCTCTGCTTTGTTTCAGACTGGAGTGCAGTGTGTGATCATGTGCTCATTTGCAAC 644  
QY 269 TTCCATTTCCAGAC 283  
DB 645 TTGAACCTCTGGGC 659

RESULT 7  
US-10-239-676-53  
;; Sequence 53, Application US/10239676  
;; Publication No. US20030082609A1  
;; GENERAL INFORMATION:  
;; APPLICANT: OLEK, Alexander  
;; APPLICANT: PIEPENROCK, Christian  
;; APPLICANT: BERLIN, Kurt  
;; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
;; FILE REFERENCE: 5013.1003  
;; CURRENT APPLICATION NUMBER: US/10/239,676  
;; CURRENT FILING DATE: 2002-09-24  
;; PRIOR APPLICATION NUMBER: PCT/EP01/03968  
;; DE 10019058.8  
;; DE 10019173.8  
;; DE 10032529.7  
;; DE 10043826.1  
;; PRIOR FILING DATE: 2001-04-06  
;; 2000-04-06  
;; 2000-04-07  
;; 2000-06-30  
;; 2000-09-01  
;; NUMBER OF SEQ ID NOS: 228  
;; SEQ ID NO 53  
;; LENGTH: 17421  
;; TYPE: DNA  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (515, 1325, 1345, 3705, 3707, 5320, 5410, 5417, 5420, 5426)  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (7028, 7082, 7179, 8249, 8252, 8254, 8269, 8275, 8278, 8279)  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (9222, 9233, 9261, 12655, 12669, 12792, 13398, 14469, 15562, 15563)

TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
 FROM VARIOUS CDNA LIBRARIES  
 FILE REFERENCE: 20411-756  
 CURRENT APPLICATION NUMBER: US/09/918, 995  
 CURRENT FILING DATE: 2001-07-30

OY	522	CCTGCAATTCCTCTGTGTGAAGTTTCTCCTTCGGAGAACAAGATTGCCTTAG	586
Db	2962	CTGACTTTCCTCCAAAGATTGCCAATGATAATCTTTGAATTCCTCCAGACACACTTAAG	2902
OY	587	GCGGAGGGAAGTCTTTTCACAGGTTCCCTCCCTTTATCTTTGTGTTTTTTTGAC	646
Db	2902	TTTGAGAGCATTTGAGTGTTGTTCTTTCTTTTCTTTTTTTTTTTTTTTTTTTCACA	2842
OY	647	CATGGGGGTTAA	658
Db	2842	AAGACGTGGTTAA	2831

RESULT 5  
US-09-892-877-103  
Sequence 103, Application US/09892877  
Publication No. US20030077809A1  
GENERAL INFORMATION:  
APPLICANT: Ruben et. al.



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:21:05 ; Search time 161 seconds  
(without alignments)  
6570.974 Million cell updates/sec

Title: US-09-477-082-1

Perfect score: 670

Sequence: 1 aagcgctcccaagcacgatt.....gggttaataaagcgttt 670

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1105431 seqs, 789497651 residues  
Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34.6	5.2	2145	US-10-003-132-12	Sequence 12, Appl
2	34	5.1	7148	US-10-239-676-8	Sequence 8, Appl
3	32.8	4.9	570	US-09-018-995-12139	Sequence 12139, A
4	32.8	4.9	7010	US-09-764-891-6031	Sequence 6031, Ap
5	32.6	4.9	704	US-09-892-877-103	Sequence 103, App
6	32.6	4.9	704	US-09-948-783-103	Sequence 103, App
7	32.6	4.9	17421	US-10-239-676-53	Sequence 53, Appl
8	32.4	4.8	1405	US-08-755-235-3	Sequence 3, Appl
9	32.4	4.8	1405	US-08-905-709-3	Sequence 3, Appl
10	32.2	4.8	1426	US-09-822-830A-512	Sequence 512, App
11	32.2	4.8	6093	US-09-764-891-8090	Sequence 8090, Ap
12	32.2	4.8	6271	US-10-172-086-35	Sequence 35, Appl
13	32.2	4.8	118951	US-10-161-572-11	Sequence 11, Appl
14	32	4.8	13606	US-10-239-676-166	Sequence 166, App
15	31.8	4.7	434	US-09-960-352-3525	Sequence 3525, Ap
16	31.8	4.7	497	US-10-198-846-1473	Sequence 1473, Ap
17	31.8	4.7	12425	US-09-876-740-50	Sequence 50, Appl
18	31.8	4.7	12425	US-10-023-529-50	Sequence 50, Appl
19	31.8	4.7	12425	US-10-023-523-50	Sequence 50, Appl

20	31.8	4.7	22635	US-09-764-891-9891	Sequence 9891, Ap
21	31.8	4.7	42999	US-09-799-462A-17	Sequence 17, Appl
22	31.8	4.7	42999	US-10-125-767-17	Sequence 17, Appl
23	31.8	4.7	42999	US-09-836-911A-17	Sequence 17, Appl
24	31.8	4.7	42999	US-10-151-081-17	Sequence 17, Appl
25	31.8	4.7	42999	US-10-287-313-17	Sequence 17, Appl
26	31.8	4.7	42999	US-10-219-694-17	Sequence 17, Appl
27	31.6	4.7	1113	US-10-028-072-109	Sequence 109, App
28	31.6	4.7	1113	US-10-121-049-109	Sequence 109, App
29	31.6	4.7	1113	US-10-123-904-109	Sequence 109, App
30	31.6	4.7	1113	US-10-140-470-109	Sequence 109, App
31	31.6	4.7	1113	US-10-175-746-109	Sequence 109, App
32	31.6	4.7	1113	US-10-176-918-109	Sequence 109, App
33	31.6	4.7	1113	US-10-176-921-109	Sequence 109, App
34	31.6	4.7	1113	US-10-137-865-109	Sequence 109, App
35	31.6	4.7	1113	US-10-140-474-109	Sequence 109, App
36	31.6	4.7	1113	US-10-142-431-109	Sequence 109, App
37	31.6	4.7	1113	US-10-143-114-109	Sequence 109, App
38	31.6	4.7	1113	US-10-140-002-109	Sequence 109, App
39	31.6	4.7	1113	US-10-142-419-109	Sequence 109, App
40	31.6	4.7	1113	US-10-123-262-109	Sequence 109, App
41	31.6	4.7	1113	US-10-142-423-109	Sequence 109, App
42	31.6	4.7	1113	US-10-121-050-109	Sequence 109, App
43	31.6	4.7	1113	US-10-141-755-109	Sequence 109, App
44	31.6	4.7	1113	US-10-143-032-109	Sequence 109, App
45	31.6	4.7	1113	US-10-123-108-109	Sequence 109, App

## ALIGNMENTS

RESULT 1  
US-10-003-132-12/c  
: Sequence 12, Application US/10003132  
: Publication No. US20020192750A1  
: GENERAL INFORMATION:  
: APPLICANT: Fox, Brian A.  
: APPLICANT: Gao, Zeren  
: APPLICANT: Shoemaker, Kimberly E.  
: TITLE OF INVENTION: NEUROFILIN HOMOLOG 2CUB5  
: FILE REFERENCE: 00-62  
: CURRENT APPLICATION NUMBER: US/10/003.132  
: PRIOR APPLICATION NUMBER: US/10-11-15  
: CURRENT FILING DATE: 2001-11-15  
: PRIOR FILING DATE: 2000-11-15  
: NUMBER OF SEQ ID NOS: 19  
: SOFTWARE: FastSeq for Windows Version 3.0  
: SEQ ID NO 12  
: LENGTH: 2145  
: TYPE: DNA  
: ORGANISM: Artificial Sequence  
: FEATURE:  
: OTHER INFORMATION: degenerate nucleotide sequence  
: NAME/KEY: misc. feature  
: LOCATION: (1)-(2145)  
: OTHER INFORMATION: n = A,T,C or G  
US-10-003-132-12

Query Match 5.2%; Score 34.6; DB 9; Length 2145;

Best Local Similarity 29.4%; Pred. No. 1.4;  
Matches 82; Conservative 48; Mismatches 148; Indels 1; Gaps 1;

QY	343	GAGTAGATAAAGCTTCCAAATPAACAGATGCCAGCGCTTGAATTTGACAGT	402
DB	1721	GCTTGTGCTGCTGACATCCCTCTTGAATGATGGAATGAGGATCTGTTCTT	462
QY	403	CCATGAATGCTGCTGACATCCCTCTTGAATGATGGAATGAGGATCTGTTCTT	462
DB	1661	TTTCCNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGT	1602
QY	463	TAAACGAGAACATTTCTTGTGAGAGATCATCTCTGTTGAGAGTAAGTT	522
DB	1601	GTDATNARTTCNARTTYTGTGATCATCTTCTCTCTCTCTCTCTCTCTCTCTCT	1543

Db	442	TGAGATTGGAGAGTTCAGAAAATGAAAATGAGGTAAAGTTGTGCTGGTGAGAG	383
QY	106	CAGGCTTAAGCCAAATGACAAATGACCAAGACCACTTCCTCTTTTCTGACGATCTA	165
Db	382	ATGGGTATATACATTTATAAAGAAATATTTTTTCTTTAATTTCTCTACTATTTA	323
QY	166	CCGCAATTCAGCCACAGGGGCTGACTT	192
Db	322	TCACCATATTTTCCCACTTTTACAT	296

Search completed: July 10, 2003, 22:32:56  
 Job time : 55 secs

OS-09-360-220-1

46 TGAATTTGAGACAGSGCCAGGATGGGACTCAGCCTGAGGACGGGTGTATCCGGAG 105

```

FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/840,146
FILING DATE: 11-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: MTA-018.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 554 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-360-220-16

Query Match 4.5%; Score 30.4; DB 3; Length 554;
Best Local Similarity 55.6%; Prec.No. 1.6;
Matches 79; Conservative 0; Mismatches 61; Indels 2; Gaps 1

QY 517 AAGTTTACCCGACAGTTCCTTCCTGCTGACAGTTTCTCTTCGAGACAGATT 576
Db 290 AATTTGGCTAGAGACACACTGCTGCTGAGAGAGTGCCTCACTGACAGAACCTG--G 233
QY 577 CTGCCTTAGCGTGGAGGAAAGTGTTCACAGGTTCTCCTTTAATCTTTTGCTTT 636
Db 232 AAGGACAGACGCTGCGAGGCCAGAGTAAAGCCCTGGTCCCTCGAGATTTCCTTAACCTTC 173
QY 637 TTTTTCGACGATGGGGTTAA 658
Db 172 CTTATGGGCTCATAGGGTTCA 151

RESULT 13
US-08-840-146-1/c
Sequence No. 1, Application US/08840146
Patent No. 6037173
GENERAL INFORMATION:
APPLICANT: Glucksmann, M. Alexandra
TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS AND METHODS AND
TITLE OF INVENTION: DIAGNOSTIC ASSAYS FOR DISEASES INVOLVING TRBP
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HONG & ELLIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/840,146
FILING DATE: 11-APR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: MTA-018.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2241 base pairs

```

```

: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: U.S.A.
: ZIP: 94111
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/633,148
: FILING DATE: 16-APR-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: MURPHY ESQ., MATTHEW B.
: REGISTRATION NUMBER: 39,787
: REFERENCE/DOCKET NUMBER: 014618-005600US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 326-2400
: TELEFAX: (415) 326-2422
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 957 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-633-148-3

Query Match
Best Local Similarity 66.2%; Score 31.2; DB 2; Length 957;
Matches 45; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 182 AGGCGTACTTTACCCAGTCCGGGGAGGAGAGAGGCGTGTGACTTCAGTGC 241
Db 766 AGGGACACACATCTTCATCCAGTGTGAGAGAGAGGCGTGGGACGACTTCACAGG 707
QY 242 TGAGGTTT 249
Db 706 TCAGGGTT 699

RESULT 9
US-08-633-148-1/c
: Sequence 1, Application US/08633148
: Patent No. 5864018
: GENERAL INFORMATION:
: APPLICANT: MORSEY, MICHAEL J.
: APPLICANT: NAGASHIMA, MARIKO
: APPLICANT: HOLLANDER, DORIS A.
: TITLE OF INVENTION: ANTIBODIES TO ADVANCED GLYCOSYLATION
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TOWNSEND & TOWNSEND & CREW LLP
: STREET: TWO EMBARCADERO CENTER, 8TH FLOOR
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: U.S.A.
: ZIP: 94111
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/633,148
: FILING DATE: 16-APR-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: MURPHY ESQ., MATTHEW B.
: REGISTRATION NUMBER: 39,787
: REFERENCE/DOCKET NUMBER: 014618-005600US
```

```

: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 326-2400
: TELEFAX: (415) 326-2422
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1023 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-633-148-1

Query Match
Best Local Similarity 66.2%; Score 31.2; DB 2; Length 1023;
Matches 45; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 182 AGGCGTACTTTACCCAGTCCGGGGAGGAGAGAGGCGTGTGACTTCAGTGC 241
Db 832 AGGGACACACATCTTCATCCAGTGTGAGAGAGAGGCGTGGGACGACTTCACAGG 773
QY 242 TGAGGTTT 249
Db 772 TCAGGGTT 765

RESULT 10
US-09-457-066-42/c
: Sequence 42, Application US/09457066
: Patent No. 6432673
: GENERAL INFORMATION:
: APPLICANT: Gao, Zeren
: APPLICANT: Hart, Charles E.
: APPLICANT: Piddington, Christopher S.
: APPLICANT: Sheppard, Paul O.
: APPLICANT: Shoemaker, Kimberly E.
: APPLICANT: Gilbertson, Debra G.
: APPLICANT: West, James W.
: TITLE OF INVENTION: GROWTH FACTOR HOMOLOG ZVEGF3
: FILE REFERENCE: 98-60
: CURRENT APPLICATION NUMBER: US/09/457,066
: CURRENT FILING DATE: 1999-12-07
: NUMBER OF SEQ ID NOS: 50
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 42
: LENGTH: 3571
: TYPE: DNA
: ORGANISM: Mus musculus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1049)...(2086)
: US-09-457-066-42

Query Match
Best Local Similarity 58.9%; Score 30.8; DB 4; Length 3573;
Matches 53; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 548 AGTTTCTCTTTCTCTCGGAGACAGATTCTGCTTACGCTGAGGAGTGTTCAC 607
Db 2372 ATTGTCTCTGTGTCGAGAGACCATCTGTGCGCTCTCAGAGGAGGCTTGTG 2313
QY 608 AGGTTCTCTCTTATCTTTGTTT 637
Db 2312 ATTGACTCTCTTGTGCTGTGCTTTT 2283

RESULT 11
US-08-840-146-16/c
: Sequence 16, Application US/08840146
: Patent No. 6037173
: GENERAL INFORMATION:
: APPLICANT: Glucksmann, M. Alexandra
: TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS AND METHODS AND
: TITLE OF INVENTION: DIAGNOSTIC ASSAYS FOR DISEASES INVOLVING TRAP
```





[illegible]



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:21:05 ; Search time 51 Seconds

(Without alignments)  
4028,891 Million cell updates/sec

Title: US-09-477-082-1

Perfect score: 670

Sequence: 1 aagcgctcccaagacagcatt.....ggggtaataaagcgttt 670

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 segs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cg2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cg2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cg2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cg2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cg2\_6/ptodata/1/ina/PCUTUS.COMB.seq:\*  
6: /cg2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	71.6	10.7	2887	US-08-983-502-14	Sequence 14, Appl
2	71.6	10.7	2887	PCT-US96-10521-14	Sequence 14, Appl
3	49.6	7.4	7218	US-08-233-463-14	Sequence 14, Appl
4	33.2	5.0	1901	PCT-US93-05000-32	Sequence 32, Appl
5	33.2	5.0	2022	US-08-464-517-32	Sequence 32, Appl
6	33.2	5.0	2022	US-08-246-361A-32	Sequence 32, Appl
7	33.2	5.0	2022	US-08-463-148-3	Sequence 32, Appl
8	31.2	4.7	957	US-08-633-148-3	Sequence 3, Appl
9	31.2	4.7	1023	US-08-633-148-1	Sequence 1, Appl
10	30.8	4.6	3573	US-09-457-066-42	Sequence 16, Appl
11	30.4	4.5	554	US-08-840-146-16	Sequence 16, Appl
12	30.4	4.5	554	US-09-360-220-16	Sequence 16, Appl
13	30.4	4.5	2241	US-08-840-146-1	Sequence 1, Appl
14	30.4	4.5	2241	US-09-360-220-1	Sequence 1, Appl
15	30.2	4.5	1457	US-08-460-512-1	Sequence 1, Appl
16	29.8	4.4	18627	US-08-961-527-113	Sequence 113, App
17	29.8	4.4	40352	US-08-846-111D-15	Sequence 15, Appl
18	29.8	4.4	40352	US-09-444-077-15	Sequence 15, Appl
19	29.6	4.4	1186	US-09-222-017B-478	Sequence 478, App
20	29.6	4.4	2248	US-08-633-237-1	Sequence 1, Appl
21	29.6	4.4	2248	US-08-975-405-1	Sequence 1, Appl
22	29.6	4.4	2264	US-09-167-109-6	Sequence 6, Appl
23	29.6	4.4	7218	US-08-233-463-14	Sequence 14, Appl
24	29.4	4.4	446	US-09-222-575-93	Sequence 93, Appl
25	29.2	4.4	976	US-08-248-839C-5	Sequence 5, Appl
26	29.2	4.4	983	US-08-248-839C-1	Sequence 1, Appl
27	29.2	4.4	1155	US-08-002-024B-7	Sequence 7, Appl

C 28	29.2	4.4	1200	US-08-248-839C-7	Sequence 7, Appl
C 29	29.2	4.4	7680	PCT-US95-09819-6	Sequence 6, Appl
C 30	29.2	4.4	7705	US-08-259-569-16	Sequence 16, Appl
C 31	29.2	4.4	7705	US-08-826-885-16	Sequence 16, Appl
C 32	29.2	4.4	7705	54551582	Patent No. 545518
C 33	29.2	4.4	7803	US-08-551-356-1	Sequence 1, Appl
C 34	29.2	4.4	7803	PCT-US93-12687-1	Sequence 1, Appl
C 35	29	4.3	8791	PCT-US96-01735-5	Sequence 5, Appl
C 36	29	4.3	13737	US-09-538-414-10	Sequence 10, Appl
C 37	29	4.3	14113	US-09-223-134-1	Sequence 1, Appl
C 38	29	4.3	14113	US-08-992-801-1	Sequence 1, Appl
C 39	29	4.3	14113	US-09-223-535-1	Sequence 1, Appl
C 40	28.8	4.3	387	US-08-686-878A-52	Sequence 52, Appl
C 41	28.8	4.3	387	US-08-721-924-2	Sequence 2, Appl
C 42	28.8	4.3	387	US-09-175-928-6	Sequence 6, Appl
C 43	28.8	4.3	53526	US-08-658-136-2	Sequence 2, Appl
C 44	28.8	4.3	53577	US-08-658-136-1	Sequence 1, Appl
C 45	28.8	4.3	4411529	US-09-103-840A-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-08-983-502-14  
Sequence 14, Application US/08983502  
Patent No. 6399327  
GENERAL INFORMATION:  
APPLICANT: David WALLACH  
APPLICANT: Mark P. BOLDIN  
APPLICANT: Taty M. GONCHAROV  
APPLICANT: Yuri V. GOLTSEV  
TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS RECEPTORS  
TITLE OF INVENTION: AND OTHER PROTEINS  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Broadway and Neimark  
STREET: 419 Seventh Street N.W., Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/983,502  
FILING DATE: 16-JAN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10521  
FILING DATE: 14-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 114,615  
FILING DATE: 16-JUL-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 114,986  
FILING DATE: 17-AUG-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 115,319  
FILING DATE: 14-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 116,568  
FILING DATE: 27-DEC-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 117,932  
FILING DATE: 16-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Broadway, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: WALLACH-19  
TELECOMMUNICATION INFORMATION:



Query Match 5.8%; Score 38.8; DB 17; Length 654;  
 Best Local Similarity 50.0%; Pred. No. 24;  
 Matches 97; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 332 ATTACTAAATGAGTCAGTATTAAGCTTTCCATAAAGCATGTCACGGCTGGCTTT 391  
 DB 322 AATATACATCTGATATTAAGAGAGTCTGAAAAGTGAATCCATCATGACAGCTTT 263  
 QY 392 AGTTTGCAGTCGACATGTTCTGACACATCCCTCTTGATGAGTGAATGGCAT 451  
 DB 262 ATTCAAGACTTTTACTAATCTGTCTAAGTGGCAGATTAATCACTAGAGTGGCATTT 203  
 QY 452 CTCCTGCTTTAAACAGGAACATTTCTGTCGAGTGCATCTCTGTCCTTTA 511  
 DB 202 TGGTTTCTTTTAAACAATCTGCATTTCTTATCTCTCTAGATTTTCTTTTCTTTT 143  
 QY 512 GGAGTAAAGTTTAC 525  
 DB 142 GCTGAATTAATTAAC 129

RESULT 13  
 AG043660/c 1080 bp DNA linear GSS 01-NOV-2001  
 LOCUS Pan troglodytes DNA, clone: PTB-022A16.R, genomic survey sequence.  
 DEFINITION AG043660  
 ACCESSION AG043660.1 GI:16572385  
 VERSION GSS.  
 KEYWORDS Pan troglodytes male lymphoblast DNA, clone\_11b:PTB Chimpanzee Male  
 SOURCE BAC Library clone:PTB-022A16.R.  
 ORGANISM Pan troglodytes  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.  
 REFERENCE 1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
 Totoki, Y., Watanabe, H. and Sakaki, Y.  
 TITLE BAC end sequences of library PTB  
 JOURNAL Unpublished  
 AUTHORS 2 (bases 1 to 1080)  
 JOURNAL Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
 Totoki, Y., Watanabe, H. and Sakaki, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan  
 1-7-22 Suhei-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:chimpesgsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,  
 Tel:81-45-503-9111, Fax:81-45-503-9170)  
 COMMENT Clones are derived from the chimpanzee BAC library PTB this BAC end  
 was generated during the Rad process and may have higher chance of  
 clone tracking errors.  
 PRIMER

Sequencing: M13Rev  
 LIBRARY  
 Vector : pKS145  
 R.Site 1 : SacI  
 R.Site 2 : SacI  
 Location/Qualifiers  
 1.1080  
 /organism="Pan troglodytes"  
 /db\_xref="taxon:9598"  
 /clone="PTB-022A16.R"  
 /sex="male"  
 /cell\_type="lymphoblast"  
 /clone\_lib="PTB Chimpanzee Male BAC Library"  
 BASE COUNT 573 a 55 c 287 g 67 t 98 others  
 ORIGIN

Query Match 5.8%; Score 38.6; DB 17; Length 1080;  
 Best Local Similarity 49.4%; Pred. No. 26;  
 Matches 83; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 474 CATTCTTGTGAGTGAATCTCTGTTCTGCTTAGAGATAAGTTACCTGACGT 533

DB 809 CTTTCCCT 750  
 QY 534 TCTTCTGTGTGAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 553  
 DB 749 TTTCT 690  
 QY 594 GGAAGTGTTCACAGGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 641  
 DB 689 TCTTCT 642

RESULT 14  
 CNS00LJ/c 1101 bp DNA linear GSS 03-JUN-1999  
 LOCUS Drosophila melanogaster genome survey sequence 17 end of BAC:  
 DEFINITION BACR30P10 of RPc1-98 library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.  
 ACCESSION AL068307  
 VERSION AL068307.1 GI:4958538  
 KEYWORDS GSS.  
 SOURCE Drosophila melanogaster.  
 ORGANISM Drosophila melanogaster.  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 1101)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 Evry cedex - FRANCE (E-mail : seque@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 COMMENT Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the Berkeley Drosophila Genome Project (BDGP).  
 The BDGP is constructing a physical map of the Drosophila  
 melanogaster genome using these BACs. For further information  
 please see <http://www.fruitfly.org/TheBDGP/Drosophila>  
 melanogaster BAC library was prepared by Kazuo Oosawa and  
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of  
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
 NY. The library is named RPc1-98 and was constructed by partial  
 EcoRI digestion of Drosophila DNA provided by the BDGP from the  
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
 P1 and EST libraries. A more detailed description of the library  
 and how to order individual BAC clones, the entire library, or  
 filters for hybridization from the BACPAC Resource Center can be  
 found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
 source  
 1.1101  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone="BACR30P10"  
 /clone\_lib="RPc1-98"  
 /note="end : 17"  
 BASE COUNT 507 a 148 c 112 g 171 t 163 others  
 ORIGIN

Query Match 5.8%; Score 38.6; DB 17; Length 1101;  
 Best Local Similarity 35.0%; Pred. No. 26;  
 Matches 97; Conservative 42; Mismatches 188; Indels 0; Gaps 0;

QY 365 AATAAGCATGTCAGGCGCTGAGTTAGTGCACGCGCATGATGTCGCCACATCC 424  
 DB 607 AATAAGCATGTCAGGCGCTGAGTTAGTGCACGCGCATGATGTCGCCACATCC 548  
 QY 425 CTTCTGCAATGTTGGAATGGGCACTCTCTCTTAACAGGAACATTTCTGTT 484  
 DB 547 CTTTWTAAAGGGGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 488  
 QY 485 CGAGTATCATCT 544  
 DB 487 TTTCTTTCT 428

[illegible][illegible]



RESULT 6	BO941968	940 bp	mRNA	linear	EST 21-NOV-2002
LOCUS	BO941968				
DEFINITION	AGNCOURT_8835770 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6421760				
ACCESSION	BO941968				
VERSION	BO941968.1	GI:22357446			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eunaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 940)				
TITLE	NIH-MGC <a href="http://imgc.ncl.nih.gov/">http://imgc.ncl.nih.gov/</a> .				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Tissue Procurement: DCPD/DTF/Gazdar cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU) DNA sequencing by: Agencourt Bioscience Corporation clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMU at: <a href="http://image.lmu.gov">http://image.lmu.gov</a> Plate: LCM2597 row: k column: 09 High quality sequence start: 13 High quality sequence stop: 625. Location/Qualifiers				
FEATURES	1..940				
SOURCE	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:6421760" /clone_lib="NIH_MGC_18" /tissue_type="large cell carcinoma" /lab_host="DH10B (phage-resistant)" /note="Organ: Lung; Vector: pCRB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by Oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCCAGGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."				
BASE COUNT	281 a 200 c 236 g 223 t				
ORIGIN					
Query Match	6.3%; Score 42.4; DB 14; Length 940;				
Best Local Similarity	89.1%; Pred. No.2.9;				
Matches	57; Conservative 0; Mismatches 6; Indels 1; Gaps 1;				
OY	555 TCCTTCCTCGGAGACGAGATTCGCCCTTACGCTGGAGGAGAGTGTTCACAGTCT 614				
DB	1 TCCTTCCTCGGA-ACCGAGATTCGCCCTTCGCTGGAGGAGAGTGTTCACAGATTA 59				
OY	615 CCTC 618				
DB	60 ATTC 63				
LOCUS	CNS04NC0/C	921 bp	DNA	linear	GSS 21-MAY-2000
DEFINITION	Tetradodon nigroviridis genome survey sequence PUC-ORI end of clone 122009 of library G from Tetradodon nigroviridis, genomic survey sequence.				
ACCESSION	AL298521	GI:8037101			
VERSION	AL298521.1				
KEYWORDS	GSS: genome survey sequence.				
SOURCE	Tetradodon nigroviridis.				
ORGANISM	Tetradodon nigroviridis				

[illegible]

BASE COUNT	253 a	179 c	182 g	184 t	123 others
Query Match					
Best Local Similarity		6.2%	Score 41.6;	DB 17;	Length 921;
Matches	71; Conservative	38; Mismatches	87; Indels	0; Gaps	0;
QY	475	ATTTCCTTTTGAGAGGAGTCACTCTGTCTTCGCTTTAGAGAGTAAGTTTACCCTGCATTT	534		
Db	483	AGTTCCTGCTTWTGCGWTTACTGTTTTTCTTKKKKTTWTGTTTTTAAWKKRAACKGYKATTC	424		
QY	535	CCCTCTGTGGTGAAGTTTCTCTCTCTCTCGAGACAGATTCTGCTTTACGCTGGAGAG	594		
Db	423	CKRTTCTGTTAKKCTTTTTTTTTTTTGGKKRTTAAKKRTWTGCTTTTTTTTTTTTWWKG	364		
QY	595	GAACTGTTTTACAGAGTTCTCCTCTTTATCTTTTGATTTTTTTTTCGACCATATGSGGG	654		
Db	363	CCWTKRTTGTGTTKMGTTTTTGMKKRYTWMCCAKRTTCTGTATVATGATTCGTCGACATGKCTC	304		
QY	655	TTAATAATAAGCGCTTT	670		
Db	303	YTTATACAMAYGCKWT	288		
RESULT 8					
CNS009VU					
LOCUS					
DEFINITION	CNS009VU	792 bp	DNA	linear	GSS 03-JUN-1998
ACCESSION	Drosophila melanogaster genome survey sequence TE13 end of BAC #				
VERSION	BACR20D15 of RPCI-98 library from Drosophila melanogaster (fruit				
KEYWORDS	fly), genomic survey sequence.				
SOURCE	AL053987				
ORGANISM	AL053987.1	GI:4935574			
REFERENCE	Drosophila melanogaster.				
	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
	Ephydroidea; Drosophilidae; Drosophila.				
	1 (bases 1 to 792)				

```

Db      62 GAAGTTTCTCTCTTCGCGAGACCAATTCCTGCTGCGAGGAAATGTTTC 121
Oy      606 ACAGGTCTCTC 618
Db      122 ACAGATTATATTC 134

RESULT 4
BI824550      804 bp      mRNA      linear      EST 04-OCT-2001
LOCUS      603035156F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5176548 5',
DEFINITION      mRNA sequence.
ACCESSION      BI824550
VERSION      BI824550.1 GI:15936100
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE      1 (bases 1 to 804)
JOURNAL      NIH-MGC http://mgc.nci.nih.gov/
COMMENT      National Institutes of Health, Mammalian Gene Collection (MGC)
              Unpublished (1999)
              Contact: Robert Strausberg, Ph.D.
              Email: cga@bbs.remall.nih.gov
              Tissue Procurement: Life Technologies, Inc.
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LMA11439 row: 0 column: 13
              High quality sequence stop: 620.
              Location/Qualifiers
                1..804
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone_image="5176548"
                  /clone_lib="NIH_MGC_115"
                  /lab_host="DH10B"
                  /note="Organ: pooled brain, lung, testis; Vector:
                  pCMV-SPORT6; Site: 1: Not; Site 2: EcoRV (destroyed); RNA
                  source anonymous pool of 6 male brains, age range 23-27; 1
                  male lung, age 27, and 1 male testis, age 65. Library is
                  oligo-dT primed and directionally cloned (EcoRV site is
                  destroyed upon cloning). Average insert size 1.8 kb,
                  insert size range 1-3 kb. Library is normalized and
                  enriched for full-length clones and was constructed by C.
                  Gruber (Invitrogen). Research Genetics tracking code
                  021. Note: this is a NIH-MGC library."
BASE COUNT      226 a      187 c      209 g      182 t
ORIGIN
Query Match      14.1%; Score 94.8; DB 13; Length 804;
Best Local Similarity 86.6%; Pred. No. 1.9e-13;
Matches 116; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

Oy      518 AAGTTACCCCTCAGTCTCTCTGCGGAGGTTTCTCTCTCTGCGAG-ACCAATT 576
Db      3 AAGTTACCTGACATCTCTCATCTGCGAAGTTTCTCTCTCTCTCGAGAACCAATT 62
Oy      577 CTGCTTACGCTGAGGAGGAGTGTTCACAGGTTCTCCTTTATCTTTGTGTTT 636
Db      63 CTGCTTCTCTGCTGAGGAGGAGTGTATTCACAGGTTCTCTCTTTATCTTTAGTGT 122
Oy      637 TTTTTCGAGCCATG 650
Db      123 TTGTTCAAGCCCTG 136

RESULT 5

```

```

CNS0176S      1101 bp      DNA      linear      GSS 26-JUL-1999
LOCUS      CNS0176S
DEFINITION      Drosophila melanogaster genome survey sequence SP6 end of 3AC
              BACN1704 of DrosBAC library from Drosophila melanogaster (fruit
              fly), genomic survey sequence.
ACCESSION      AL108022
VERSION      AL108022.1 GI:5628326
KEYWORDS      GSS.
SOURCE      Drosophila melanogaster.
ORGANISM      Drosophila melanogaster.
REFERENCE      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
AUTHORS      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
TITLE      Ephyroidea; Drosophilidae; Drosophila.
JOURNAL      1 (bases 1 to 1101)
              Genoscope.
              Direct Submission
              Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
              BP 191 91006 EVRY cedex - FRANCE (E-mail: seqret@genoscope.cns.fr
              - Web : www.genoscope.cns.fr)
              Determination of this BAC-end sequence was carried out as part of a
              collaboration with the European Drosophila Genome Project (EDGP) -
              http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
              library (Dros BAC) was made by Alain Billard at CEPH (Centre
              d'Etude du Polymorphisme Humain) with funding provided by a MGC
              project grant. The DNA was prepared from embryos by Alain Bucheton
              and Genevieve Payan. It has been constructed in the vector
              pBelosBAC11.
              Location/Qualifiers
                1..1101
                  /organism="Drosophila melanogaster"
                  /db_xref="taxon:7227"
                  /clone="BACN1704"
                  /clone_lib="DrosBAC"
                  /plasmid="pBelosBAC11"
                  /note="end : SP6"
BASE COUNT      268 a      126 c      100 g      195 t      412 others
ORIGIN
Query Match      7.28; Score 48.4; DB 17; Length 1101;
Best Local Similarity 21.9%; Pred. No. 0.089;
Matches 98; Conservative 144; Mismatches 205; Indels 0; Gaps 0;

Oy      222 CTGCTGCTGACTTCAGTCTGCTGAGGTTGATCAAGCAAGGAGAACTCTTCCTCCAG 281
Db      594 YTTTCTTCTGTTTWTWMTTWTAKMCTTYARGGANMWTTHSCGACSCMNATYHC 653
Oy      282 ACCCTTGCAGAAAGAAATGACATATTACTTGCCTGCGGAGGGGTTATTACTAAT 341
Db      654 AAHTWRAVMDCBANWTTHTAABAABGSTTKAAANAYAMKTGTAAACAANAWA 713
Oy      342 GGAATCAGTAAATAGCTTCCACATAAAGATGTCAGGCTGGGCTGAGTTTGCAGC 401
Db      714 TYWMMYATTTKYIYTAHMHMAAAMCYCSYCKSGTSCBSTRKYGTTCGTG 773
Oy      402 TCCATGATTTCTGCACACATCCCTCTCTGAAATGATGATGAGATGCGATCTGTCT 461
Db      774 SCSTYVAAANDSSSYTOSYCSYTTTKTKTKTKTKTKGKGBSBSBTYTTTTYTT 833
Oy      462 TTAACACGGAACATTTCTTCTGAGTCAATCTCTCTCTCTCTCTTCTTGGAGTAAAG 521
Db      834 TYCBBBGGBBKGTCTCTTTCYCCYCCYBBSYBHYCYCYBYCCYCTYKDKMRAWTTS 893
Oy      522 TTACCTGCAATTCCTCTGCTGAGGAGTGTTCCTCTCTCTGAGAGACCAAGATTCGCC 581
Db      894 TKBRKTKTKTKTKTKTKSKGKKKKKTSBKIVBYGIBYGGCISBSCKTKTYBC 953
Oy      582 TTACGCTGAGGAGAGTGTTCACAGGTTCTCCTCTCTTATCTTTGTTGTTTTTTT 641
Db      954 YCCYTTCSIKTTTKTKYKTBTTTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 1013
Oy      642 CGACCCATGGGGCTTAATAAAGCCT 668
Db      1014 BBYKSSGYSCSKBSBCKBYGCBCT 1040

```

FEATURES  
source

Location/Qualifiers  
1. 522  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZp313f1234"  
/clone\_1lb="313 (synonym: hlc3)"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="vector: pTriplex2; Site\_1: SflIA; Site\_2: SflIB;  
CDNA-collection"

BASE COUNT 120 a 116 c 133 g 149 t 4 others

ORIGIN

Query Match 22.6%; Score 151.6; DB 9; Length 522;  
Best Local Similarity 96.2%; Pred. No. 1e-27; Indels 0; Gaps 0;  
Matches 154; Conservative 0; Mismatches 6;

QY 491 AGTCATCTCTGTCCTTGGAGTAAGTTTACCTGACGTTCTCTGTGTGAAGT 550  
|||||  
DB 1 AGTCATCTCTGTCCTTGGAGTAAGTTTACCTGACGTTCTCTGTGTGAAGT 60  
|||||

QY 551 TTCTCTTCTCTGCGAGACCAAGTTCGCTTACGCTGAGGAGAGTGTTCACAGG 610  
|||||  
DB 61 TTCTCTTCTCTGCGAGACCAAGTTCGCTTACGCTGAGGAGAGTGTTCACAGG 120  
|||||

QY 611 TTCTCTCTCTTTATCTTTGTTGTTTTCGAGCCATG 650  
|||||  
DB 121 TTCTCTCTCTTTATCTTTGTTGTTTTCGAGCCCTG 160  
|||||

RESULT 2  
AL702116 421 bp mRNA linear EST 22-MAR-2002  
LOCUS DKFZp686E18155\_r1 686 (synonym: hlc3) Homo sapiens CDNA clone  
DEFINITION DKFZp686E18155 5', mRNA sequence.  
ACCESSION AL702116  
VERSION AL702116.1 GI:19685471  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 421)  
Bloeker, H., Boecher, M., Brandt, P., Mewes, W., Well, B. and Wiemann  
S. (2002) Bloeker, H., Boecher, M., Brandt, P., Mewes, H.W., Well, B. and  
Wiemann, S.)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Bloeker H  
MIPS  
Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de,  
sequenced by GBR (National Research Centre for Biotechnology Ltd.,  
Braunschweig/Germany) within the CDNA sequencing consortium of the  
German Genome Project.  
No sl sequence available.  
This clone (DKFZp686E18155) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
Location/Qualifiers  
1. 421  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZp686E18155"  
/clone\_1lb="586 (synonym: hlc3)"  
/tissue\_type="human skeletal muscle"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="vector: pTriplex2; Site\_1: SflIA; Site\_2: SflIB;  
CDNA-collection"

BASE COUNT 95 a 89 c 113 g 124 t

ORIGIN

Query Match 19.7%; Score 132; DB 9; Length 421;  
Best Local Similarity 96.4%; Pred. No. 8.9e-23; Indels 0; Gaps 0;  
Matches 135; Conservative 0; Mismatches 5;

QY 511 AGGAGTAAGTTTACCCTGAGTTCCTTCTGCTGAGAGTTTCTTCTCGGAGAC 570  
|||||  
DB 2 AGGAGTAAGTTTACCCTGAGTTCCTTCTGCTGAGAGTTTCTTCTCGGAGAC 61  
|||||

QY 571 CAGATTCCTGCTTACCTGAGGAGAGTTTTCACAGGTTCTCTCTTATCTTT 630  
|||||  
DB 62 CAGATTCCTGCTTACCTGAGGAGAGTTTTCACAGGTTCTCTCTTATCTTT 121  
|||||

QY 631 GTGTTTTTTTCGAGCCATG 650  
|||||  
DB 122 GTGTTTTTTTCGAGCCCTG 141  
|||||

RESULT 3  
AL702289 463 bp mRNA linear EST 22-MAR-2002  
LOCUS DKFZp686M12156\_r1 686 (synonym: hlc3) Homo sapiens CDNA clone  
DEFINITION DKFZp686M12156 5', mRNA sequence.  
ACCESSION AL702289  
VERSION AL702289.1 GI:19685644  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 463)  
Bloeker, H., Boecher, M., Brandt, P., Mewes, W., Well, B. and Wiemann  
S. (2002) Bloeker, H., Boecher, M., Brandt, P., Mewes, H.W., Well, B. and  
Wiemann, S.)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Bloeker H  
MIPS  
Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de,  
sequenced by GBR (National Research Centre for Biotechnology Ltd.,  
Braunschweig/Germany) within the CDNA sequencing consortium of the  
German Genome Project.  
No sl sequence available.  
This clone (DKFZp686M12156) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
Location/Qualifiers  
1. 463  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZp686M12156"  
/clone\_1lb="586 (synonym: hlc3)"  
/tissue\_type="human skeletal muscle"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="vector: pTriplex2; Site\_1: SflIA; Site\_2: SflIB;  
CDNA-collection"

BASE COUNT 127 a 98 c 110 g 128 t

ORIGIN

Query Match 18.2%; Score 121.8; DB 9; Length 463;  
Best Local Similarity 94.7%; Pred. No. 3.2e-20; Indels 0; Gaps 0;  
Matches 126; Conservative 0; Mismatches 7;

QY 486 GAGTGAATCATCTGTTGCTTACGAGTAAGTTTACCTGAGATTCCTTCTGTGT 545  
|||||  
DB 2 GAGTGAATCATCTGTTGCTTACGAGTAAGTTTACCTGAGATTCCTTCTGTGT 61  
|||||

QY 546 GAAATTTCTCTCTCTGAGACAGATTTCGCTTACGCTGAGGAGAGTGTTCG 605  
|||||



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 20:26:44 ; Search time 1349 Seconds  
(without alignments)  
8043.718 Million cell updates/sec

Title: US-09-477-082-1

Perfect score: 670  
Sequence: 1 aagcgctccaagacacgctt.....ggggttaataaagcgctt 670

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estinu:\*  
5: em\_estrov:\*  
6: em\_estrpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_estf2:\*  
11: gb\_hic:\*  
12: gb\_estc3:\*  
13: gb\_estf4:\*  
14: gb\_estf5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pin:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	151.6	22.6	9	AL600571	AL600571 DKFZ313F
2	132	19.7	421	AL702116	AL702116 DKFZP686E
3	121.8	18.2	463	AL702289	AL702289 DKFZP686E
4	94.8	14.1	804	B1824550	B1824550 603035156
5	48.4	7.2	1101	CNS0176S	AL108022 Drosophila
6	42.4	6.3	940	BQ941968	BQ941968 AGENCOURT

Result No.	Score	Query Match	Length DB	ID	Description
7	41.6	6.2	921	CNS04NC0	AL298521 Tetraodon
8	41.4	6.2	792	CNS009YU	AL053997 Drosophila
9	39.8	5.9	784	B1857533	B1857533 603384318
10	39.8	5.8	398	T86125	T86125 yd48405.s1
11	38.8	5.8	423	CNS001W9	AL075520 Drosophila
12	38.8	5.8	654	AQ0528592	AQ0528592 RPI-11-3
13	38.6	5.8	1080	AG043660	AG043660 Pan trogl
14	38.6	5.8	1101	CNS00LJT	AL068307 Drosophila
15	38.4	5.7	805	CNS011W	AL100646 Drosophila
16	38.2	5.7	377	BF109017	BF109017 714906.x
17	38.2	5.7	514	BQ023350	BQ023350 UT-1-BBip
18	38	5.7	916	CNS0126V	AL101188 Drosophila
19	38	5.7	1182	BM562140	BM562140 AGENCOURT
20	37.6	5.6	256	BG467316	BG467316 LA14F09 B
21	37.6	5.6	838	CNS03W9T	AL263440 Tetraodon
22	37.6	5.6	907	BM415809	BM415809 OP20891 M
23	37.4	5.6	304	BH256367	BH256367 LDH10MCRO
24	37.4	5.6	496	B1448281	B1448281 dac74f05.
25	37.4	5.6	518	AL515195	AL515195 AL515195
26	37.4	5.6	559	AZ711718	AZ711718 RPI-24-1
27	37.2	5.6	378	BE069215	BE069215 OV3-BT037
28	37.2	5.6	567	AV748759	AV748759 AV748759
29	37.2	5.6	627	AZ632355	AZ632355 1M0486M23
30	37	5.5	1151	B1077468	B1077468 602872286
31	37	5.5	1201	CNS016FD	AL106675 Drosophila
32	37	5.5	1424	BG166139	BG166139 602340748
33	36.8	5.5	298	B1055127	B1055127 PMO-GN034
34	36.8	5.5	503	AI397303	AI397303 fb09f07.x
35	36.8	5.5	750	CNS011HD	AL100303 Drosophila
36	36.8	5.5	763	AG072907	AG072907 Pan trogl
37	36.8	5.5	1101	CNS00LKL	AL068436 Drosophila
38	36.8	5.5	1101	CNS00LYT	AL078921 Drosophila
39	36.6	5.5	728	BH259496	BH259496 CH230-157
40	36.6	5.5	1263	BE730099	BE730099 601563260
41	36.4	5.4	272	BF702042	BF702042 MI-P-A1-a
42	36.4	5.4	402	BM336284	BM336284 KEST192-A
43	36.4	5.4	883	BG032662	BG032662 602298302
44	36.4	5.4	925	BF668274	BF668274 60212471
45	36.4	5.4	964	BG034757	BG034757 602325401

## ALIGNMENTS

RESULT 1  
AL600571 522 bp mRNA linear EST 14-AUG-2001  
DERIVATION DKFZP313F1234.F1 313 (synonym: hloc2) Homo sapiens cDNA clone  
ACCESSION DKFZP313F1234.5', mRNA sequence.  
VERSION AL600571.1 GI:15164077  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 522)  
Ansoerge,W., Witzner,U., Mewes,W., Weill,B. and Wiemann,S.  
EST (Ansoerge,W., Witzner,U., Mewes,H.W., Weill,B. and Wiemann,S.)  
Unpublished (1999)  
COMMENT Contact: Ansoerge W  
MIPS  
Am Klopferstr. 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;  
Heidelberg/Germany) within the cDNA sequencing consortium of the  
German Genome Project.  
No st sequence available.  
This clone (DKFZP313F1234) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.





[illegible]

PR	26-JUL-2000	2000US-02209564
PR	26-JUL-2000	2000US-02209564
PR	16-AUG-2000	2000US-02243438
PR	14-AUG-2000	2000US-02245113
PR	14-AUG-2000	2000US-02252519
PR	14-AUG-2000	2000US-02252519
PR	14-AUG-2000	2000US-02252516
PR	14-AUG-2000	2000US-02252566
PR	14-AUG-2000	2000US-02252567
PR	14-AUG-2000	2000US-02252567
PR	14-AUG-2000	2000US-02252568
PR	14-AUG-2000	2000US-02252570
PR	14-AUG-2000	2000US-02254547
PR	14-AUG-2000	2000US-02257577
PR	14-AUG-2000	2000US-02257578
PR	14-AUG-2000	2000US-02257598
PR	18-AUG-2000	2000US-02265779
PR	22-AUG-2000	2000US-02266681
PR	22-AUG-2000	2000US-02266681
PR	22-AUG-2000	2000US-02271782
PR	22-AUG-2000	2000US-02271782
PR	30-SEP-2000	2000US-02289282
PR	30-SEP-2000	2000US-02289282
PR	01-SEP-2000	2000US-02293433
PR	01-SEP-2000	2000US-02293433
PR	01-SEP-2000	2000US-02293445
PR	05-SEP-2000	2000US-02295503
PR	05-SEP-2000	2000US-02295503
PR	06-SEP-2000	2000US-02304337
PR	06-SEP-2000	2000US-02304337
PR	08-SEP-2000	2000US-02311422
PR	08-SEP-2000	2000US-02311423
PR	08-SEP-2000	2000US-02311444
PR	08-SEP-2000	2000US-02314133
PR	08-SEP-2000	2000US-02314143
PR	08-SEP-2000	2000US-02320801
PR	12-SEP-2000	2000US-02313668
PR	14-SEP-2000	2000US-02323977
PR	14-SEP-2000	2000US-02323988
PR	14-SEP-2000	2000US-02323989
PR	14-SEP-2000	2000US-02324000
PR	14-SEP-2000	2000US-02324001
PR	14-SEP-2000	2000US-02323603
PR	14-SEP-2000	2000US-02323604
PR	14-SEP-2000	2000US-02330655
PR	21-SEP-2000	2000US-02344223
PR	21-SEP-2000	2000US-02344274
PR	25-SEP-2000	2000US-02349977
PR	25-SEP-2000	2000US-02349988
PR	25-SEP-2000	2000US-02354844
PR	27-SEP-2000	2000US-02358344
PR	27-SEP-2000	2000US-02358346
PR	29-SEP-2000	2000US-02363367
PR	29-SEP-2000	2000US-02363368
PR	29-SEP-2000	2000US-02363369
PR	02-OCT-2000	2000US-02368002
PR	02-OCT-2000	2000US-02370337
PR	02-OCT-2000	2000US-02370338
PR	02-OCT-2000	2000US-02370339
PR	02-OCT-2000	2000US-02370400
PR	13-OCT-2000	2000US-02399355
PR	13-OCT-2000	2000US-02399357
PR	20-OCT-2000	2000US-02409560
PR	20-OCT-2000	2000US-02411221
PR	20-OCT-2000	2000US-02411785
PR	20-OCT-2000	2000US-02411786
PR	20-OCT-2000	2000US-02411787
PR	20-OCT-2000	2000US-02418068
PR	20-OCT-2000	2000US-02418069
PR	01-NOV-2000	2000US-02418565
PR	08-NOV-2000	2000US-02464674
PR	08-NOV-2000	2000US-02464674

PR 13-DEC-2000; 2000US-255281P.  
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 PA  
 XX Schlegel R, Endege WO, Monahan JE;  
 PI  
 DR WPI; 2001-662795/76.  
 XX  
 PT Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer  
 XX  
 PS Claim 1; Page 4923-4924; 11750pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for:  
 CC (a) assessing whether a patient is afflicted with prostate cancer;  
 CC (b) monitoring the progression of prostate cancer in a patient;  
 CC (c) assessing the efficacy of a test compound to inhibit prostate  
 CC cancer in a patient;  
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
 CC in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound;  
 CC (g) determining whether prostate cancer has metastasized in a patient;  
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
 CC patient;  
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
 XX  
 SO Sequence 2819 BP; 787 A; 651 C; 571 G; 809 T; 1 other;  
 XX  
 Query Match 5.3%; Score 35.4; DB 23; Length 2819;  
 Best Local Similarity 51.6%; Pred. No. 3.1;  
 Matches 81; Conservative 0; Mismatches 76; Indels 0; Gaps 0;  
 QY 498 TCTGTTCTCTTGAAGAGTAAGTTACCGCTCAGTCTCTGCGTGAAGTTTCTCT 557  
 Db 97 TCTGTTCTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 156  
 QY 558 TTTCTCTGGAGACACAGATTCTGCTTAAGCTGAGGAGAGTGTTCACAGAGTTCTCT 617  
 Db 157 CGGTCTGATGTCGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 216  
 QY 618 CCTTTTATCTTTGTTGTTTCTTTTTCGAGGATGGGG 654  
 Db 217 CGTCACTATCTCTGTTTCTGCCCCGCTCCGGCG 253  
 XX  
 RESULT 12  
 AAF98635/C  
 ID AAF98635 standard; DNA; 301 BP.  
 XX  
 AC AAF98635;  
 XX  
 DT 02-JUL-2001 (first entry)  
 XX  
 DE Human ovarian cancer cell expressed sequence 6285.  
 XX  
 KM Human; ovarian cancer; identification; detection; characterisation;  
 KM tumour; kinase; marker; cytostatic; antisense gene therapy; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200118542-A2.  
 PD 15-MAR-2001.  
 XX  
 PF 01-SEP-2000; 2000WO-US24199.  
 XX  
 PR 03-SEP-1993; 99US-0152547.  
 PR 16-MAR-2000; 2000US-0190347.  
 PR 21-MAR-2000; 2000US-0191321.

PR 31-MAY-2000; 2000US-0208382.  
 PR 20-JUL-2000; 2000US-0220467.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Lee J, Thompson P, Lillie J;  
 DR WPI; 2001-211428/21.  
 XX  
 PT Detection, assessment, prevention and therapy of ovarian cancer,  
 PT comprises detecting changes in the expression of a variety of markers  
 XX  
 PS Claim 1; Page 991; 1198pp; English.  
 XX  
 CC The present invention describes a method for assessing whether a patient  
 CC is afflicted with ovarian cancer by comparing: (1) the expression of a  
 CC marker (I) (see AAF98594 to AAF98730), in a patient sample; and (2) the  
 CC normal level of expression of (I) in a control non-ovarian cancer  
 CC sample, where a significant difference between the level of expression  
 CC in (a) and (b) is an indication that the patient is afflicted with  
 CC ovarian cancer. (I) have cytostatic activities and can be used in  
 CC antisense gene therapy. The method, compositions and kits from the  
 CC present invention can be used for: (1) assessing and treating ovarian  
 CC cancer; (2) making isolated hybridoma, which produces an antibody useful  
 CC for ovarian cancer assessment; and (3) inhibiting ovarian cancer in a  
 CC patient. AAF98573 to AAF98593 represent human kinase marker primers and  
 CC probes which are used in the exemplification of the present invention.  
 XX  
 SO Sequence 301 BP; 52 A; 86 C; 63 G; 97 T; 3 other;  
 XX  
 Query Match 5.3%; Score 35.2; DB 22; Length 301;  
 Best Local Similarity 54.7%; Pred. No. 1.3;  
 Matches 70; Conservative 0; Mismatches 58; Indels 0; Gaps 0;  
 QY 210 GGGAGGAGAGCGCTGCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 269  
 Db 299 GGGAGGCGAAGGCGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240  
 QY 270 TCTATTCCAGACCCCTTTCGAGAAAGATGCAATTAATCTGCGCGACAGGGGTTA 329  
 Db 239 GGGCAAAACCTGCTCTATGATAAAATAGTAGTCTGCTGCTGCTGCTGCTGCTGCT 180  
 QY 330 TTATTACT 337  
 Db 179 GTATCACT 172  
 XX  
 RESULT 13  
 AAQ53209  
 ID AAQ53209 standard; DNA; 2020 BP.  
 XX  
 AC AAQ53209;  
 XX  
 DT 22-JUN-1994 (first entry)  
 XX  
 DE Human cyclin D3 pseudogene.  
 XX  
 KM D-type; mammalian; CLN protein; protein deficiency; cell cycle start;  
 KM yeast; complement; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FT Key Location/Qualifiers  
 FT CDS 1136..1793  
 FT /\*tag= a  
 XX  
 PN WO9324514-A.  
 PD 09-DEC-1993.  
 XX  
 PF 25-MAY-1993; 93WO-US05000.  
 XX  
 PR 26-MAY-1992; 92US-0888178.



CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

XX Sequence 916 BP; 138 A; 93 C; 321 G; 364 T; 0 other;

SO Query Match 5.4%; Score 36.2; DB 24; Length 916;

Best Local Similarity 50.9%; Pred. No. 1;  
Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 475 ATTCTCTTGAGTGAATCATCTCTCTGCTTTAGAGTAAGTTACCCGCACTT 534

DB 326 ATTTCGGTTTAAATTCGGCGTTTCGTTGAATTCGTTTCTGTTTGTATTATA 385

OY 535 CCTTCGTGTGAGTGTCTCTCTCTCTCGAGACAGATTTCGCTTACGCTGAGG 594

DB 386 TTTTCTTTTCTGTTTCTGTTTCTTTTAACTTTTAAAGTCGGCGGCTTTGCGGTGCA 445

OY 595 GAAGTGTTCACAGGTTCTCCTCTTTATCTTTGTGTTTTCG 643

DB 446 GAGGTTTACGCTTCGTTTTCGTTTCTGTTTCTTTTATTTTCGTTTCG 494

RESULT 8

ABQ49947/c

ID ABQ49947 standard; DNA; 916 BP.

XX ABQ49947;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36538.

XX Human; cytosine methylation; 5'-CPG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

KW SNP; cell differentiation; ds.

XX Homo sapiens.

PF WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

DR WPI; 2002-371829/40.

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of

CC methylation of a particular cytosine in a motif 5'-CPG-3', present in a

CC genomic sample of DNA. The sample is treated chemically to convert

CC cytosine (C) but not methylated C, to uracil, then part of the genomic

CC DNA that contains the target C is amplified to form a labeled amplicon.

CC The amplicon is hybridised to two classes, each with at least one

CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers

CC and the degree of hybridisation to both classes is determined from the

CC label on the amplicon. From the ratio of labels hybridised to the two

CC classes of oligomers, the degree of methylation is calculated. The method

CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's) and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

XX Sequence 916 BP; 364 A; 321 C; 93 G; 138 T; 0 other;

SO Query Match 5.4%; Score 36.2; DB 24; Length 916;

Best Local Similarity 50.9%; Pred. No. 1;  
Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 475 ATTCTCTTGAGTGAATCATCTCTCTGCTTTAGAGTAAGTTACCCGCACTT 534

DB 591 ATTTCGGTTTAAATTCGGCGTTTCGTTGAATTCGTTTCTGTTTGTATTATA 532

OY 535 CCTTCGTGTGAGTGTCTCTCTCTCTCGAGACAGATTTCGCTTACGCTGAGG 594

DB 531 TTTTCTTTTCTGTTTCTGTTTCTTTTAACTTTTAAAGTCGGCGGCTTTGCGGTGCA 472

OY 595 GAAGTGTTCACAGGTTCTCCTCTTTATCTTTGTGTTTTCG 643

DB 471 GAGGTTTACGCTTCGTTTTCGTTTCTGTTTCTTTTATTTTCGTTTCG 423

RESULT 9

ABQ67119

ID ABQ67119 standard; DNA; 6431 BP.

XX ABQ67119;

DT 28-AUG-2002 (first entry)

DE Human angiogenesis associated polynucleotide SEQ ID NO 149.

XX Human; angiogenesis; methylation; eye disease; glaucoma; tumour;

KW inflammation; rheumatoid arthritis; diabetic retinopathy; anti-leukers;

KW macular degeneration; inflammatory bowel disease; Crohn's disease;

KW antihypertensive; antiarrhythmic; antidiabetic; antipsoriatic;

KW antiarteriosclerotic; ds.

XX Homo sapiens.

PF WO200246454-A2.

XX 13-JUN-2002.

XX 06-DEC-2001; 2001WO-EP14320.

PR 06-DEC-2000; 2000DE-1061338.

XX (EPIC-) EPIGENOMICS AG.

XX Schacht O;

DR WPI; 2002-500450/53.

XX Claim 1; SEQ ID NO 149; 41pp + Sequence Listing; German.

XX New nucleic acid fragments from chemically treated

PT angiogenesis-associated genes, useful for determining methylation

PT status, e.g. in diagnosis or treatment of cancer

XX The invention relates to a nucleic acid (I) comprising a segment of 18

CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)

CC having sequences (ABQ66971-ABQ67176) or their complements. (I), also

CC related oligomers, are used to evaluate the methylation status and/or

CC single-nucleotide polymorphisms, in angiogenesis-related genes, for

ABK84723  
 ID ABK84723 standard; CDNA; 2887 BP.  
 AC ABK84723;  
 XX  
 DT 14-AUG-2002 (first entry)  
 DE Human cDNA differentially expressed in granulocytic cells #1294.  
 XX  
 KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
 KW viral infection; parasitic infection; protozoal infection;  
 KW fungal infection; sterile inflammatory disease; psoriasis;  
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
 KW adult respiratory distress syndrome; inflammatory bowel disease;  
 KW Crohn's disease; ulcerative colitis; periodontal disease;  
 KW granulocyte activation; chronic inflammation; allergy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200228999-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PE 03-OCT-2001; 2001WO-US30821.  
 XX  
 PR 03-OCT-2000; 2000US-237189P.  
 XX  
 PA (GENE-) GENE LOGIC INC.  
 XX  
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
 DR WPI: 2002-435328/46.  
 XX  
 PT Detecting granulocyte activation by detecting differential expression  
 PT of genes associated with granulocyte activation, which serves as  
 PT diagnostic markers that is useful for monitoring disease states and  
 PT drug toxicity -  
 XX  
 PS Claim 1; SEQ ID NO 1294; 114pp; English.  
 XX  
 CC The invention relates to detecting (M1) granulocyte (GC) activation  
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
 CC DNA chip analysis as given in the specification, and comparing  
 CC the expression level to an expression level in an unactivated  
 CC GC, where differential expression of Gs is indicative of GCA.  
 CC Also included are modulating (M2) GA by contacting Gs with an agent  
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)  
 CC for an agent capable of modulating GCA or an inflammation (especially  
 CC chronic) in a tissue, an allergic response in a subject, exposure of a  
 CC subject to a pathogen or sterile inflammatory disease using the  
 CC gene expression profile; (3) detecting (M4) an inflammation (especially  
 CC chronic) in a tissue, an allergic response in a subject, exposure of a  
 CC subject to a pathogen or sterile inflammatory disease, by detecting the  
 CC level of expression in a sample of the tissue of gene(s) from Gs, where  
 CC the level of expression of the gene is indicative of inflammation;  
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,  
 CC an allergic response in a subject, exposure of a subject to a pathogen  
 CC or sterile inflammatory disease, by contacting a tissue having  
 CC inflammation with an agent that modulates the expression of gene(s)  
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
 CC modulating Gs; M3 is useful for screening an agent capable of modulating  
 CC GCA preferably in an inflammation in a tissue; M4 is useful for  
 CC detecting an inflammation (especially chronic) in a tissue, an allergic  
 CC response in a subject, exposure of a subject to a pathogen or sterile  
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,  
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal  
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,  
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,  
 CC periodontal disease; also bacterial infection, viral infection,  
 CC parasitic infection, protozoal infection, fungal infection and M5 is  
 CC useful for treating one of the above conditions. The present  
 CC sequence represents a gene differentially expressed in granulocytes.

CC Note: the sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pot\_sequences.  
 CC  
 XX

SQ Sequence 2887 BP; 853 A; 595 C; 625 G; 814 T; 0 other;

Query Match 10.7%; Score 71.6; DB 24; Length 2887;  
 Best Local Similarity 94.9%; Pred. No. 2.2e-11;  
 Matches 74; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 573 GATTTCGCTTACGTCGAGGAGGAACTGTTTCACAGGTTTCCTTTATCTTTGT 632

DB 1 GATTTCGCTTTCGTCGAGGAGGAACTGTTTCACAGGTTTCCTTTATCTTTGT 60

QY 633 GTTTTTCGAGCCATG 650

DB 61 GTTTTTCAGCCCTG 78

RESULT 7

ABQ49946  
 ID ABQ49946 standard; DNA; 916 BP.

AC ABQ49946;  
 XX

DT 12-JUL-2002 (first entry)  
 XX

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36537.  
 XX

KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX

OS Homo sapiens.  
 XX

PN WO200218632-A2.  
 XX

PD 07-MAR-2002.  
 XX

PE 01-SEP-2001; 2001WO-EP10074.  
 XX

PR 01-SEP-2000; 2000DE-1043826.  
 XX

PR 05-SEP-2000; 2000DE-1044543.  
 XX

PA (EPIC-) EPIGENOMICS AG.  
 XX

PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX

DR WPI: 2002-371829/40.  
 XX

PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 XX  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue



/product= MORT-1 binding protein (MACH) isoform alpha

XX MO9703998-A1.  
 XX 06-FEB-1997.  
 XX 14-JUN-1996; 96WO-US10521.  
 XX 16-APR-1996; 96IL-0117932.  
 XX 16-JUL-1995; 95IL-0114615.  
 XX 17-AUG-1995; 95IL-0114986.  
 XX 14-SEP-1995; 95IL-0115319.  
 XX 27-SEP-1995; 95IL-0116588.  
 XX (WEIN/) WEINURZEL H.  
 XX (YEDA) YEDA RES & DEV CO LTD.  
 XX Boidin M, Goltssev YV, Goncharov T, Wallach D;  
 XX WPI: 1997-132570/12.  
 XX P-PSDB: AAM11891.  
 XX New DNA encoding MACH protein that interacts with MORT-1 protein -  
 PT to mediate intracellular effects of FAS or TNF receptors, partic.  
 PT for regulating apoptosis in tumours, virus-infected cells etc.  
 PS Example 3: Page 116-117; 163pp; English.  
 XX AAT61396, and AAT61405-T61411 represent coding sequences for different  
 CC isoforms of MACH. MACH is a binding protein for the mediator of receptor  
 CC toxicity (MORT-1) protein. MORT-1 binds to the FAS ligand receptor  
 CC (FAS-R) death domain region, and triggers part of the cell death  
 CC signalling cascade in mammalian cells. Vectors containing these  
 CC sequences, the encoded proteins, and antibodies (Ab) against them are  
 CC used to modulate the effect of FAS-R ligand or TNF on cells that carry  
 CC FAS-R or p55-R. This is specifically for treating tumours, HIV-infected  
 CC cells or other diseased cells, by control of apoptosis/programmed cell  
 CC death. The encoded proteins are mediators of the cell death pathway  
 CC initiated by TNF and FAS-R binding, i.e. it mimics or enhances the  
 CC effect of MORT-1 where increased cytotoxicity is required. To inhibit the  
 CC effect of MORT-1, e.g. in cases of septic shock, graft rejection and  
 CC acute hepatitis, sequences encoding antisense molecules or ribozymes, or  
 CC Ab against the protein, are used. Compounds that inhibit MACH are  
 CC potentially useful for controlling MACH activity e.g. in cases of  
 CC autoimmune disease, oligodendrocyte death in multiple sclerosis or  
 CC AIDS-inhibited T-cell suicide. The MACH proteins can also be used to  
 CC isolate and characterise other proteins and receptors involved in  
 CC signalling and for Ab production. The Ab can be used to purify the new  
 CC proteins and for diagnosis of conditions involving abnormal function of  
 CC FAS-R mediated cellular effects.  
 XX Sequence 2887 BP; 853 A; 595 C; 625 G; 814 T; 0 other;  
 SO Query Match 10.7%; Score 71.6; DB 18; Length 2887;  
 Best Local Similarity 94.9%; Pred. No. 2.2e-11;  
 Matches 74; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 573 GATCTGCCTTTACGCTGGAGGAGGAGTGTTCACAGGTTCTCTTATCTTTGT 632  
 DB 1 GATCTGCCTTTACGCTGGAGGAGGAGTGTTCACAGGTTCTCTTATCTTTGT 60  
 OY 633 GTTTTTCGAGCATG 650  
 DB 61 GTTTTTCAGGCTTG 78

RESULT 5  
 ID AAD03915 standard; DNA; 2887 BP.  
 XX AAD03915;  
 XX 02-JUL-2001 (first entry)

XX Human caspase-8 DNA.  
 DE Human; caspase-8; MACH; FLICE; Mch3;  
 XX cysteine protease; apoptosis; caspase expression cassette; metastasis;  
 KW tumour; cathepsin B; urokinase; proliferation; gene therapy; ds.  
 XX Homo sapiens.  
 XX Key Location/Qualifiers  
 FT CDS 292..1731  
 FT /tag= a  
 FT /product= "Human caspase-8"  
 XX MO200129232-A2.  
 XX 26-APR-2001.  
 XX 19-OCT-2000; 2000WO-US28941.  
 XX 20-OCT-1999; 99US-0160559.  
 XX 14-AUG-2000; 2000US-0225564.  
 XX (SCIO-) SCIOS INC.  
 XX Cordell B, Li Y;  
 XX WPI: 2001-290920/30.  
 XX P-PSDB; AAE00605.  
 XX Novel fusion polypeptide comprising first and second caspase subunit  
 PT separated by cleavage site not associated in nature with caspase  
 PT subunit, useful for cloning gene encoding enzymes involved in  
 XX proteolytic cleavage -  
 XX Disclosure; Fig 15; 116pp; English.  
 XX The present sequence is a DNA encoding human Caspase-8 also known as  
 CC MACH, FLICE and Mch3. Caspases are a family of cysteine proteases, that  
 CC participate in the initiation and execution of apoptosis. Caspases exist  
 CC as pro-enzymes, activated by cleavage into a large and small subunit,  
 CC occurring after specific aspartic acid residues within the pro-enzyme  
 CC sequence. The present invention relates to a method for functional  
 CC cloning of genes encoding proteins or enzymes involved in proteolytic  
 CC cleavage. The invention is based on the use of caspase expression  
 CC cassettes comprising the coding sequence of a proteolytic cleavage site  
 CC flanked by sequences encoding two caspase subunits. A fusion polypeptide  
 CC comprising a first and a second caspase subunit, separated by a cleavage  
 CC site not associated in nature, is useful for cloning gene encoding  
 CC enzymes involved in proteolytic cleavage. An expression cassette  
 CC containing fusion polypeptide is used to identify a mutant cell line  
 CC deficient in an enzyme of interest and is also useful for diagnosis and  
 CC suppression of proliferation or metastases of a tumour cell characterised  
 CC by overexpression of a polypeptide (e.g. Cathepsin B or urokinase,  
 CC selectively expressed in the tumour cells). DNA encoding fusion  
 CC polypeptide is used in gene therapy.  
 XX Sequence 2887 BP; 853 A; 595 C; 625 G; 814 T; 0 other;  
 SO Query Match 10.7%; Score 71.6; DB 22; Length 2887;  
 Best Local Similarity 94.9%; Pred. No. 2.2e-11;  
 Matches 74; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 573 GATCTGCCTTTACGCTGGAGGAGGAGTGTTCACAGGTTCTCTTATCTTTGT 632  
 DB 1 GATCTGCCTTTACGCTGGAGGAGGAGTGTTCACAGGTTCTCTTATCTTTGT 60  
 OY 633 GTTTTTCGAGCATG 650  
 DB 61 GTTTTTCAGGCTTG 78

RESULT 6

XX	(HUMA-) HUMAN GENOME SCI INC.
PA	Rosen CA, Barash SC, Ruben SM;
XX	WPI: 2001-483426/52.
DR	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT	useful for preventing, diagnosing and/or treating cancers and
PT	metastasis -
XX	Disclosure; SEQ ID NO 32029; 3071bp + Sequence Listing; English.
PS	AAs54951 to AAs64702 encode the human immune/hematopoietic antigen (I)
CC	amino acid sequences given in AAs62170 to AAs91921. (I) have cytosolic
CC	activity, and can be used in gene therapy and vaccine production. (II)
CC	proteins and polynucleotides may be used in the prevention, diagnosis and
CC	treatment of diseases associated with inappropriate (I) expression. For
CC	example, they may be used to treat disorders associated with decreased
CC	expression by rectifying mutations or deletions in a patient's genome
CC	that affect the activity of (I) by expressing inactive proteins or to
CC	suppress the patients own production of (I). Additionally, (I)
CC	polynucleotides may be used to produce the secreted (I), by inserting
CC	the nucleic acids into a host cell and culturing the cell to express the
CC	protein. (I) proteins and polynucleotides may be used to prevent,
CC	diagnose and treat immune/hematopoietic-related diseases, especially
CC	cancers and cancer metastases of hematopoietic-derived cells. AAs64703
CC	to AAs87594 represent human immune/hematopoietic antigen genomic
CC	sequences from the present invention. AAs54942 to AAs54950 and AAs82169
XX	represent sequences used in the exemplification of the present invention.

Seq	Sequence	45017 BP;	12919 A;	10037 C;	9566 G;	12495 T;	0 other:							
Query Match	77.08;	Score 516;	DB 22;	Length 45017;										
Best Local Similarity	94.45;	Pred. No. 7,6e-148;												
Matches	568;	Conservative	0;	Mismatches	30;	Indels	4;							
						Gaps	3;							
Oy	AATTGGAAGCAAGCGCCAGATGGAGATCGACCGCTGAGCAGGGGTGATCCGGAGCA 108													
Db	26162 AAATGGAGAAACATGAGCCAAAGATGGAACTCAGCCTGAGCAGGGCTGAGCC--GAGCAG 26105													
Oy	109 GCGTAAAGCAAGTACGATGAACACAGCACACTTCCTCTCTTTTTCGAAAGATACCC 168													
Db	26104 GGGTAAACCAAGTGGAGGCTGAGCAGAGCACTTCCTCTCTTTT-TCGAAACATCTACCT 26046													
Oy	169 GCATTAGACCCAGAGGGCTGACTTTAGCCAGTCCGGCGGGAGGAGAGAGGGCTGTCT 228													
Db	26045 ACATTTAGCTACAGGGGCTGGCTTTACCCAGTCCGGCGGGAGGAGAGAGGGCTGTCT 25986													
Oy	229 GTGACTTCAGTGCAGAGGTTTGATCAAGGCAAGGAGAACTTCCTATCCAGACCTTT 288													
Db	25985 GTGACTTCAGTGCAGAGGTTTGATCAAGGCAAGGAGAACTTCCTATCCAGACCTTT 25922													
Oy	289 GCAGAAAGAAATGCGATATTACTTTGCCGCCGACAGGAGGTTATTATTACTAAATGAGTCA 348													
Db	25925 GCAAGAAAGAAATGCGATATTACTTTGCCGCCGACAGGAGGTTATTATTACTAAATGAGTCA 25866													
Oy	349 GTATAATGCTTCCATTAAGAGATGCCAGCGCTGGGCTTAGTTGCAGCTCCATCA 408													
Db	25865 GTATAATGCTTCCATTAAGAGATGCCAGCGCTGGGCTTAGTTGCAGCTCCATCA 25806													
Oy	409 ATTGTGCGCCACATCCCTCTTCGTGAATGTTGGAAATTTGGGCATCTGTTCCTTAAACA 468													
Db	25805 ATTGTGCGCCACATCCCTCTTCGTGAATGTTGGAAATTTGG-GCATGTGTTCCCTTAAACA 25744													
Oy	469 GGAACACTTCTGTGTGTGAGTGAATCATCTCTGTTTGTGAGTAAAGTTACCT 528													
Db	25746 GGAACACTTCTGTGTGTGAGTGAATCATCTCTGTTTGTGAGTAAAGTTACCT 25687													
Oy	529 GCACTTCCTCTGAGTGAATTTTCCTTCTCTGAGACAGATTCGCTTTACCC 588													
Db	25686 GCACTTCCTCTGAGTGAATTTTCCTTCTCTGAGACAGATTCGCTTTACCC 25622													
Oy	589 TGGAGGGAAGTGTTCACAGGTTCTCTCTCTTTATATCTTTGTGTTTTTCGAGCA 648													
Db	25626 TGGAGGGAAGTGTTCACAGGTTCTCTCTCTTTATATCTTTGTGTTTTTCGAGCCC 25567													
Oy	649 TGG 650													
Db	25566 TGG 25565													
RESULT 4														
AAAT61405														
AAAT61405 standard; cDNA; 2887 BP.														
AC	AAAT61405;													
XX	29-OCT-1997 (first entry)													
DE	MACH isoform alpha1 coding sequence.													
XX	MACH; MORF-1 binding protein; mediator of receptor toxicity; cell death;													
KW	antibody; FAS ligand receptor; FAS-R; death domain region; septic shock;													
KW	tumour necrosis factor; tumour; HIV-infection; oligodendrocyte death;													
KW	apoptosis/programmed cell death; p53-R; graft rejection; acute hepatitis;													
KW	autoimmune disease; multiple sclerosis; AIDS-inhibited T-cell suicide;													
XX	TNF; therapy; ss.													
OS	Homo sapiens.													
XX	Key													
TH	Location/Qualifiers													
FT	291..1731													

Db 24906 GTGACTCAGTGTGAGGTTTGATCAAGCAAGGAACCTCCATTCACACCCCTT 24847  
QY 289 GCAGAGAAAGATGGCATATTACTTGGCCGACAGAGGGTATTATTAATGAGTCA 348  
Db 24846 GCAGAGAAAGATGGCATATTACTTGGCCGACAGAGGGTATTATTAATGAGTCA 24787  
QY 349 GATTAATGTTTCCATAAAGCATGTCCAGCCCTGGGGCTTGTAGTTCACGTCATGA 408  
Db 24786 GTATAATGTTTCCATAAAGCATGTCCAGCCCTGGGGCTTGTAGTTCACGTCATGA 24727  
QY 409 ATTGTCTGCACATCCCTCTTCGAAATGTTGGAATGGGCATCTGTCTTAAACA 468  
Db 24726 ATTGTCTGCACATCCCTCTTCGAAATGTTGGAATGGGCATCTGTCTTAAACA 24668  
QY 469 GGAACATTTCTGTGAGTGAAGTATCTGTGCTTGGAGTAAAGTTACCT 528  
Db 24667 GGAACATTTCTGTGAGTGAAGTATCTGTGCTTGGAGTAAAGTTACCT 24608  
QY 529 GCAATTCCTCTGTGAGTGAAGTATCTGTGCTTGGAGTAAAGTTACCT 588  
Db 24607 GCAATTCCTCTGTGAGTGAAGTATCTGTGCTTGGAGTAAAGTTACCT 24548  
QY 589 TGGAGGAAAGTGTTCACAGGTTTCCTCTTATCTTTGTTTTCGAGCA 648  
Db 24547 TGGAGGAAAGTGTTCACAGGTTTCCTCTTATCTTTGTTTTCGAGCC 24488  
QY 649 TG 650  
Db 24487 TG 24486

RESULT 3  
AAK77217/c  
ID AAK77217 standard; DNA: 45017 BP.

XX AAK77217:  
AC 07-NOV-2001 (first entry)  
DT 07-NOV-2001 (first entry)  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32029.  
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32029.  
KW Human: immune; haematopoietic; immune/haematopoietic antigen; cancer:  
XX cytostatic; gene therapy; vaccine; metastasis; ds.  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX 09-AUG-2001.  
PD 17-JAN-2001; 2001MO-US01354.  
PE 17-JAN-2001; 2001MO-US01354.  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198133.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226881.  
PR 22-AUG-2000; 2000US-0226888.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239835.  
PR 13-OCT-2000; 2000US-0239837.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.

```

PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231247.
PR 08-SEP-2000; 2000US-0231247.
PR 08-SEP-2000; 2000US-0231247.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231411.
PR 08-SEP-2000; 2000US-0231411.
PR 08-SEP-2000; 2000US-0231411.
PR 08-SEP-2000; 2000US-0230801.
PR 08-SEP-2000; 2000US-0230801.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0231397.
PR 14-SEP-2000; 2000US-0231398.
PR 14-SEP-2000; 2000US-0231399.
PR 14-SEP-2000; 2000US-0231400.
PR 14-SEP-2000; 2000US-0231401.
PR 14-SEP-2000; 2000US-0231063.
PR 14-SEP-2000; 2000US-0231064.
PR 14-SEP-2000; 2000US-0231065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0238935.
PR 13-OCT-2000; 2000US-0238937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.

PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 03-DEC-2000; 2000US-0251030.
PR 03-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX (HMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis.
XX
XX Disclosure: SEQ ID NO 32028; 3071bp + sequence listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 43938 BP; 12772 A; 9656 C; 9206 G; 12304 T; 0 other;

Query Match 77.0%; Score 516; DB 22; Length 43938;
Best Local Similarity 94.4%; Pred. No. 7.5e-148;
Matches 568; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 49 AATTGAGAGACAGGCGCAAGATGGAGATGAGACCTGAGACGCGTTGATCCGAGCAG 108
DB 25083 AAATGGAGACATGGCGCAAGATGGAGATGAGACCTGAGAGCGCGAGCC--GAGCAG 25026
QY 109 GGCTAAGCAGATGAGATGAGATGAGACGACACTTCTCTTTTTCCTGAGACGATCAACC 168
DB 25025 GGCTAAGCAGATGAGGCTGAGCGAGCAACACTTCTCTTTT--TCGAAACAATCAACCT 24967
QY 169 GCATTTCAGCAGAGGCGTCACTTACCCAGTCCGGGAGGAGAGAGAGAGGCGTGCT 228
DB 24966 ACATTTCAGCAGAGGCGTCACTTACCCAGTCCGGGAGGAGAGAGAGAGGCGTGCT 24907
QY 229 GTGACTTCAGTGTGAGGCTTGTATCAAGGCAAGGAACTCTCTATTCACAGCCCTTT 288

```

Qy 541 GTGGTGAAGCTTTTCTCTTTCTCTCGAGACAGATTCTGCCCTTTACGCTGAGGGGAAGTG 6000  
|||||

PR 01-SEP-2000; 2000US-0229345.

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 20:49:45 ; Search time 223 Seconds  
(without alignments)  
6766.094 Million cell updates/sec

Title: US-09-477-082-1

Perfect score: 670  
Sequence: 1 aagcgtcccaagacagatt.....ggggttaataaagcgttt 670

Scoring table: IDENTITY-NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N.Geneseq.101002.\*  
1: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	670	100.0	670	21	AAAS1792
2	516	77.0	43938	22	AAK77216
3	516	77.0	45017	22	AAK77217
4	71.6	10.7	2887	18	AAAT61405
5	71.6	10.7	2887	22	AAAD03915
6	71.6	10.7	2887	24	ABK84723
7	36.2	5.4	916	24	ABQ49946
8	36.2	5.4	916	24	ABQ49947
9	36	5.4	6431	24	ABQ67119

10	36	5.4	6432	22	AAAS6557	Tumour suppressor
11	35.4	5.3	2819	23	ABV25186	Human prostate exp
12	35.2	5.3	301	22	AAFP8635	Human ovarian canc
13	35.2	5.3	2020	14	AAQ53209	Human cyclin D3 ps
14	35	5.2	256	22	AAK81668	Human immune/haema
15	35	5.2	6741	21	AAH10595	Gene encoding a su
16	35	5.2	19734	24	ABL33932	Human immune syste
17	34.6	5.2	581	24	ABQ15812	Oligonucleotide fo
18	34.6	5.2	581	24	ABQ15813	Oligonucleotide fo
19	34.2	5.1	13573	24	ABL33869	Human immune syste
20	34	5.1	631	24	ABQ25138	Oligonucleotide fo
21	34	5.1	631	24	ABQ25139	Oligonucleotide fo
22	34	5.1	5518	24	ABL54318	Chemically treated
23	34	5.1	5518	24	ABN80019	Human immune syste
24	34	5.1	5518	24	ABL32217	Human immune syste
25	34	5.1	7148	22	AAAS45303	Chemically pretrea
26	34	5.1	7148	24	ABN80035	Human chemically m
27	34	5.1	7148	24	ABK28142	DNA transcription
28	34	5.1	13249	24	ABL70131	Chemically treated
29	34	5.1	13249	24	ABL32116	Human immune syste
30	34	5.1	13249	24	ABK31176	Signal transductio
31	33.8	5.0	5860	24	ABL70523	Chemically treated
32	33.8	5.0	5860	24	AAAS61372	Human gene regulat
33	33.8	5.0	5860	24	ABK31456	Signal transductio
34	33.8	5.0	16891	24	ABN80058	Human chemically m
35	33.8	5.0	16891	24	ABL32652	Human immune syste
36	33.6	5.0	16891	24	ABL34504	Human metastasis a
37	33.6	5.0	1992	21	AAAC33992	Arabidopsis thalia
38	33.6	5.0	5371	24	ABL34295	Human immune syste
39	33.6	5.0	6104	24	ABL33124	Human immune syste
40	33.6	5.0	6577	22	AAAS46718	Tumour suppressor
41	33.2	5.0	2022	13	AAQ31877	Cyclin D3 pseudoge
42	33.2	5.0	26747	24	ABL67784	Nucleotide sequenc
43	33	4.9	6833	24	ABL33145	Human immune syste
44	33	4.9	8897	23	ABL20304	Drosophila melanog
45	33	4.9	16167	24	ABL70254	Chemically treated

## ALIGNMENTS

RESULT 1	AAAS1792	standard; DNA: 670 BP.
AAAS1792	31-OCT-2000 (first entry)	
XX	CASP8 promoter Region 1.	
XX	CASP8; caspase-8; promoter; Region 1; inactivation; methylation;	
XX	Cysteine protease; tumour suppressor; chromosome 2q33-34; neuroblastoma;	
XX	cancer; death receptor; apoptosis; cytosolic; gene therapy; ss.	
XX	Homo sapiens.	
XX	WO200039347-A1.	
XX	06-JUL-2000.	
XX	30-DEC-1999; 99WO-US31280.	
XX	31-DEC-1998; 98US-0114308.	
XX	(STUD-) ST JUDE CHILDREN'S RES HOSPITAL.	
XX	Kidd VJ, Lahti JM, Teitz T;	
XX	WPI; 2000-45242/39.	
XX	Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or	
XX	prognosing cancer, comprises detecting a modification of genomic DNA	

/note="L1MA10 repeat: matches 5479. .6315 of consensus"  
L1MA10 repeat: matches 5479. .6315 of consensus"  
repeat\_region  
12979. .13263  
/note="AlusX repeat: matches 9. .290 of consensus"  
AlusX repeat: matches 9. .290 of consensus"  
13548. .13778  
/note="L1M4 repeat: matches 15. .262 of consensus"  
L1M4 repeat: matches 15. .262 of consensus"  
14035. .14188  
/note="MER5A repeat: matches 43. .189 of consensus"  
MER5A repeat: matches 43. .189 of consensus"  
14785. .14859  
/note="L1ME3A repeat: matches 5785. .5859 of consensus"  
L1ME3A repeat: matches 5785. .5859 of consensus"  
15050. .15303  
/note="L1ME3A repeat: matches 5525. .5777 of consensus"  
L1ME3A repeat: matches 5525. .5777 of consensus"  
15317. .15774  
/note="L1M4 repeat: matches 3951. .4454 of consensus"  
L1M4 repeat: matches 3951. .4454 of consensus"  
15337. .15774  
/note="L1 repeat: matches 3951. .4433 of consensus"  
L1 repeat: matches 3951. .4433 of consensus"  
15943. .16022  
/note="L2 repeat: matches 2664. .2750 of consensus"  
L2 repeat: matches 2664. .2750 of consensus"  
/note="L2 repeat: matches 2664. .2750 of consensus"  
complement(16038. .16616)  
/note="match: GSS: Em:A0541664"  
17138. .17193  
/note="MIR repeat: matches 78. .133 of consensus"  
MIR repeat: matches 78. .133 of consensus"  
18611. .18904  
/note="MLTIC repeat: matches 193. .466 of consensus"  
MLTIC repeat: matches 193. .466 of consensus"  
18905. .19230  
/note="Alusg repeat: matches 1. .310 of consensus"  
Alusg repeat: matches 1. .310 of consensus"  
19231. .19419  
/note="MLTIC repeat: matches 1. .193 of consensus"  
MLTIC repeat: matches 1. .193 of consensus"  
join(19542. .19799,20379. .20477)  
/gene="dJ38088.3"  
join(<19542. .19799,20379. .>20477)  
/gene="dJ38088.3"  
/product="dJ38088.3 (a novel mRNA)"  
/note="match: ESTs: Em:R79825"  
/evidence=not\_experimental  
19669. .20180  
/gene="dJ38088.3"  
/note="match: GSS: Em:A0401011"  
20155. .20184  
/note="15 copies 2 mer ac 100% conserved"  
20482. .20806  
/note="L1R16C repeat: matches 56. .387 of consensus"  
L1R16C repeat: matches 56. .387 of consensus"  
21088. .21177  
/note="FLAM.A repeat: matches 10. .99 of consensus"  
FLAM.A repeat: matches 10. .99 of consensus"  
21469. .22143  
/note="HERVL repeat: matches 4345. .5045 of consensus"  
HERVL repeat: matches 4345. .5045 of consensus"  
21469. .21856  
/note="HERVL repeat: matches 4644. .5045 of consensus"  
HERVL repeat: matches 4644. .5045 of consensus"  
22192. .23403  
/note="HERVL repeat: matches 2953. .4189 of consensus"  
HERVL repeat: matches 2953. .4189 of consensus"  
23429. .24388  
/note="HERVL repeat: matches 1902. .2891 of consensus"  
HERVL repeat: matches 1902. .2891 of consensus"  
23611. .24388  
/note="HERVL repeat: matches 1902. .2701 of consensus"  
HERVL repeat: matches 1902. .2701 of consensus"  
24949. .25225  
/note="HERVL repeat: matches 882. .1164 of consensus"  
HERVL repeat: matches 882. .1164 of consensus"  
25786. .26094  
/note="Alu repeat: matches 1. .307 of consensus"  
Alu repeat: matches 1. .307 of consensus"  
26354. .26477  
repeat\_region

repeat\_region  
/note="L1ME repeat: matches 458. .585 of consensus"  
26716. .26792  
/note="L2 repeat: matches 2666. .2741 of consensus"  
L2 repeat: matches 2666. .2741 of consensus"  
complement(26932. .27500)  
/note="match: GSS: Em:A0417969"  
27047. .27174  
/note="MIR repeat: matches 65. .196 of consensus"  
MIR repeat: matches 65. .196 of consensus"  
27177. .27238  
/note="L2 repeat: matches 2644. .2709 of consensus"  
L2 repeat: matches 2644. .2709 of consensus"  
27243. .27405  
repeat\_region  
Query Match 6.3% Score 42.2; DB 9; Length 166882;  
Best Local Similarity 59.7%; Pred. No. 0.12; 48; Indels 0; Gaps 0;  
Matches 71; Conservative 0; Mismatches 0;  
QY 328 TATTATTACTAATGAGTCAGTATAAATGCTTCCATATAAGCATGTCAGCGCTGGG 387  
Db 163927 TATTATTTCGATATGTAATAATAAAGACCTTATAAAGCTTACCTTTTGGT 163986  
QY 388 CTTTATTGACAGTCGACGAAATGCTGCGACATGCCCTCTGTAAGTGGAAATG 446  
Db 163987 ACTAAGTTTCAAGACCGATATGATTTTACACTTCGCGCATATGCGCACTTTGGAAATG 164045

Search completed: July 11, 2003, 04:06:58  
Job time : 2000 secs

## FEATURES

Location/Qualifiers  
1..223387  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-10B21"  
/clone\_lib="RPCI mouse BAC library 23"

BASE COUNT 66248 a 45169 c 45091 g 65176 t 1703 others  
ORIGIN

Query Match 6.4%; Score 42.8; DB 2; Length 223387;  
Best Local Similarity 51.6%; Pred. No. 0.078;  
Matches 98; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 452 CTCGTCCTTAAACGAGAACATTTCTGTCGAGTGAATCTCTCTGCTTAA 511  
DB 84132 CTCCTCTCCAGACAGCCACCTATATGCTCTGACACCACTTTAGCCAC 84073  
QY 512 GGAGTAAGTTACCTCGAGTTCCTCTGTCGAGTTTCTTTCTCGAGACC 571  
DB 84072 GGACAAACTCAATCCTTGTCTACTTCCACCCCTCCTCACTGAGCTAACCC 84013  
QY 572 AGATTGCGCTTACGTCGAGGAGAGTCTTTCACAGGTTCTCTCTTTATCTTTG 631  
DB 84012 AGCCTCAGTTTCCCTTTCGAGAGGTTGCTGAGACAGCACTTTCTTTTCTTTT 83953  
QY 632 TGTCTTTT 641  
DB 83952 TTTTCTTTT 83943

## RESULT 15

AL136307 186882 bp DNA linear PRI 01-AUG-2000  
LOCUS AL136307  
DEFINITION Human DNA sequence from clone RP3-380B8 on chromosome 6p24.1-25.3  
Contains a gene encoding the protein Neuritin, which is involved in  
proliferation of neurite outgrowth, a private kinase (PKM2)  
pseudogene, a novel mRNA, 4 Cpg islands, ESTs, STS and GSSs,  
complete sequence.

ACCESSION AL136307  
VERSION AL136307.12 GI:8574124  
KEYWORDS HMG: CPG Island; Neuritin; PKM2.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 186882)  
AUTHORS Brown, J.  
TITLE Direct Submission  
JOURNAL Submitted (28-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humgen@sanger.ac.uk  
requests: clonerequests@sanger.ac.uk

## COMMENT

On Jun 20, 2000 this sequence version replaced gi:8573783.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence has been finished according to sequencing map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
EMBL: SW, SWISSPROT: Tr, TREMBL: Wp, WORMPEP: Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
Chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6

## FEATURES

source

RP3-380B8 is from the library RPCI-3 constructed at the Roswell  
Park Cancer Institute by the group of Pieter de Jong. For further  
details see <http://bacpac.med.buffalo.edu/>  
VECTOR: pCYPAC2  
This sequence is the entire insert of clone RP3-380B8 the true left  
end of clone RP11-525021 is at 141303 in this sequence. The true  
right end of clone RP1-182016 is at 57138 in this sequence.

Location/Qualifiers  
1..186882  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="p24.1-25.3"  
/clone="RP3-380B8"  
/clone\_lib="RPCI-3"  
247..465  
repeat\_region  
/note="MIR repeat: matches 16..236 of consensus  
MIR repeat: matches 16..236 of consensus"  
1442..2391  
repeat\_region  
/note="LIMB3 repeat: matches 5281..6174 of consensus  
LIMB3 repeat: matches 5281..6174 of consensus"  
2435..2499  
repeat\_region  
/note="MIR repeat: matches 77..145 of consensus  
MIR repeat: matches 77..145 of consensus"  
2540..3786  
repeat\_region  
/note="LIP2 repeat: matches 4900..6146 of consensus"  
2542..3786  
repeat\_region  
/note="LIP5 repeat: matches 4900..6143 of consensus"  
4034..4653  
repeat\_region  
/note="LIM9 repeat: matches 4022..4632 of consensus  
LIM9 repeat: matches 4022..4632 of consensus"  
4654..4953  
repeat\_region  
/note="AlusX repeat: matches 1..300 of consensus  
AlusX repeat: matches 1..300 of consensus"  
4954..6496  
repeat\_region  
/note="LIM9 repeat: matches 4632..6308 of consensus  
LIM9 repeat: matches 4632..6308 of consensus"  
6522..6820  
repeat\_region  
/note="MLTID repeat: matches 14..283 of consensus  
MLTID repeat: matches 14..283 of consensus"  
6821..6890  
repeat\_region  
/note="MER67A repeat: matches 374..543 of consensus  
MER67A repeat: matches 374..543 of consensus"  
6987..7184  
repeat\_region  
/note="MLTID repeat: matches 308..505 of consensus  
MLTID repeat: matches 308..505 of consensus"  
7290..7462  
repeat\_region  
/note="MIR repeat: matches 71..262 of consensus"  
7357..7462  
repeat\_region  
/note="MIR repeat: matches 71..177 of consensus"  
8007..8093  
repeat\_region  
/note="LIM4 repeat: matches 5688..5772 of consensus  
LIM4 repeat: matches 5688..5772 of consensus"  
8094..8329  
repeat\_region  
/note="LIM8 repeat: matches 6029..6173 of consensus  
LIM8 repeat: matches 6029..6173 of consensus"  
8932..8994  
repeat\_region  
/note="L2 repeat: matches 2637..2699 of consensus"  
8996..9199  
repeat\_region  
/note="MER33 repeat: matches -9..192 of consensus  
MER33 repeat: matches -9..192 of consensus"  
9200..9559  
repeat\_region  
/note="THE1B repeat: matches 1..364 of consensus  
THE1C repeat: matches 1..371 of consensus"  
9560..9685  
repeat\_region  
/note="MER33 repeat: matches 192..324 of consensus  
MER33 repeat: matches 192..324 of consensus"  
10364..10771  
repeat\_region  
/note="MLTIG repeat: matches 65..512 of consensus  
MLTIG repeat: matches 65..512 of consensus"  
10775..10875  
repeat\_region  
/note="MIR repeat: matches 134..252 of consensus"  
12145..12963  
repeat\_region



Query Match	6.4%;	Score 42.8;	DB 9;	Length 138006;
Best Local Similarity	45.7%;	Pred. No. 0.073;		
Matches 149;	Conservative	0;	Mismatches 177;	Indels 0;
				Gaps 0;

```

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 223387)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 223387)
DOE Joint Genome Institute.
Direct Submission
Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1740757
Center clone name: RPCI-23_10B21
-----
Summary Statistics
Consensus quality: 212195 bases at least Q40
Consensus quality: 219405 bases at least Q30
Consensus quality: 220848 bases at least Q20
Estimated insert size: 208000; agarose-1p estimation
Estimated insert size: 222587; sum-of-ctdigs estimation
Quality coverage: 10.31 in Q20 bases; agarose-1p estimation
Quality coverage: 9.63 in Q20 bases; sum-of-ctdigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 24944: contig of 24944 bp in length
24945 25044: gap of unknown length
25045 31043: contig of 5999 bp in length
31044 31143: gap of unknown length
31144 49115: contig of 17972 bp in length
49116 49215: gap of unknown length
49216 67918: contig of 18703 bp in length
67919 68019: gap of unknown length
68019 79367: contig of 11349 bp in length
79368 79467: gap of unknown length
79468 85787: contig of 6320 bp in length
85788 85887: gap of unknown length
85888 92600: contig of 6713 bp in length
92601 92700: gap of unknown length
92701 95687: contig of 2987 bp in length
95688 95787: gap of unknown length
95788 100230: contig of 4443 bp in length
100231 100330: gap of unknown length
100331 116173: contig of 15843 bp in length
116174 116273: gap of unknown length
116274 121323: contig of 5050 bp in length
121324 121433: gap of unknown length
121434 124666: contig of 3263 bp in length
124687 124786: gap of unknown length
124787 154386: contig of 29600 bp in length
154387 154486: gap of unknown length
154487 154512: contig of 9726 bp in length
154513 164312: gap of unknown length
164313 169021: contig of 4709 bp in length
169022 169121: gap of unknown length
169122 171321: contig of 2200 bp in length
171322 171421: gap of unknown length
171422 211231: contig of 39810 bp in length
211232 211331: gap of unknown length
211332 223387: contig of 12056 bp in length.

```

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL, Sw, SWISSPROT, Tr, TrEMBL, Wp, WormPep, information on the WormPep database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormPep](http://www.sanger.ac.uk/Projects/C_elegans/wormPep) This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Ch6> RPl1-294H11 is from the library RPl1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.choxi.org/Dacpac/home.htm> VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RPl1-294H11. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RPl3-448K1 is at 137907 in this sequence. The true right end of clone RPl1-600I9 is at 100 in this sequence.

## FEATURES

## SOURCE

1. 138006  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RPl1-294H11"  
/clone\_lib="RPl1-11.2"  
1. 100  
/note="match: STS: Em:HS60019B"  
2014. 2059  
/note="23 copies 2 mer tg 97% conserved"  
2149. 2446  
/note="MTR1B repeat: matches 96. 390 of consensus"  
2466. 2583  
/note="AluYb repeat: matches 1. 118 of consensus"  
4339. 4637  
/note="AluY repeat: matches 1. 298 of consensus"  
4842. 5151  
/note="AluSp repeat: matches 1. 312 of consensus"  
5152. 5332  
/note="AluSq repeat: matches 134. 312 of consensus"  
5803. 6023  
/note="AluYb repeat: matches 77. 289 of consensus"  
6056. 6364  
/note="AluY repeat: matches 1. 306 of consensus"  
complement(6200. 6688)  
/note="match: GSS: Em:AQ528743"  
complement(6225. 6677)  
/note="match: GSS: Em:AQ524727"  
complement(6414. 6688)  
/note="match: GSS: Em:AQ564844"  
6902. 7096  
/note="MER20 repeat: matches 1. 218 of consensus"  
7660. 7756  
/note="HAL1 repeat: matches 60. 161 of consensus"  
7883. 8555  
/note="LIMC4 repeat: matches 6798. 7502 of consensus"  
8556. 8860  
/note="AluSp repeat: matches 1. 308 of consensus"  
8861. 8904  
/note="LIMC4 repeat: matches 7502. 7543 of consensus"  
complement(8906. 9376)  
/note="match: GSS: Em:AQ479008"  
9175. 9469  
/note="match: GSS: Em:AQ434542"

repeat\_region 9221. 9246  
/note="13 copies 2 mer tg 96% conserved"  
repeat\_region 9258. 9466  
/note="HERV1 repeat: matches 2068. 2279 of consensus"  
repeat\_region 9482. 9677  
/note="MER45 repeat: matches 1. 178 of consensus"  
9681. 9901  
/note="HERV16 repeat: matches 1794. 2013 of consensus"  
9902. 10208  
/note="AluX repeat: matches 1. 310 of consensus"  
10209. 10314  
/note="HERV16 repeat: matches 1689. 1794 of consensus"  
10348. 10586  
/note="HERV16 repeat: matches 1357. 1577 of consensus"  
10621. 10915  
/note="AluY repeat: matches 1. 295 of consensus"  
10952. 10991  
/note="MER52A repeat: matches 1063. 1107 of consensus"  
11073. 11271  
/note="HERV16 repeat: matches 981. 1197 of consensus"  
11411. 11987  
/note="HERV16 repeat: matches 140. 813 of consensus"  
11988. 12295  
/note="AluX repeat: matches 4. 303 of consensus"  
12296. 12444  
/note="HERV16 repeat: matches 2. 140 of consensus"  
12753. 13081  
/note="AluY repeat: matches 1. 308 of consensus"  
13088. 13206  
/note="MIR repeat: matches 102. 223 of consensus"  
13335. 13611  
/note="AluSc repeat: matches 29. 305 of consensus"  
13930. 14527  
/note="LIM4A repeat: matches 5706. 6298 of consensus"  
15521. 15805  
/note="AluX repeat: matches 11. 296 of consensus"  
16061. 16130  
/note="Char1ies repeat: matches 32. 108 of consensus"  
16922. 17050  
/note="FLM1C repeat: matches 1. 143 of consensus"  
17052. 17237  
/note="AluSg/x repeat: matches 104. 274 of consensus"  
17806. 18234  
/note="LOR1B repeat: matches 32. 460 of consensus"  
18338. 18414  
/note="MTR11 repeat: matches 329. 402 of consensus"  
18415. 18720  
/note="AluYb8 repeat: matches 1. 304 of consensus"  
18721. 18966  
/note="MTR11 repeat: matches 68. 329 of consensus"  
19033. 19350  
/note="AluS repeat: matches 1. 301 of consensus"  
19791. 19897  
/note="LIMB6 repeat: matches 5096. 5205 of consensus"  
19904. 20370  
/note="LIMB6 repeat: matches 5159. 5621 of consensus"  
20372. 20760  
/note="MSTRC repeat: matches 1. 401 of consensus"  
20764. 21227  
/note="LIMB6 repeat: matches 5625. 6100 of consensus"  
21235. 21530  
/note="AluS repeat: matches 1. 396 of consensus"  
21600. 21713  
/note="MER93 repeat: matches 19. 136 of consensus"  
21714. 22019  
/note="AluS repeat: matches 1. 305 of consensus"  
22020. 22219  
/note="MER93 repeat: matches 136. 394 of consensus"  
22220. 22316  
/note="MER57B repeat: matches 296. 394 of consensus"  
22950. 23214  
/note="match: GSS: Em:B85271"  
23820. 24103  
repeat\_region

```

* 4342 4441: gap of unknown length
* 4442 6030: contig of 1589 bp in length
* 6031 6130: gap of unknown length
* 6131 7908: contig of 1778 bp in length
* 7909 8008: gap of unknown length
* 8009 9035: contig of 1027 bp in length
* 9036 9135: gap of unknown length
* 9136 10139: contig of 1004 bp in length
* 10140 10239: gap of unknown length
* 10240 11425: contig of 1186 bp in length
* 11426 11525: gap of unknown length
* 11526 13078: contig of 1553 bp in length
* 13079 13178: gap of unknown length
* 13179 14359: contig of 1181 bp in length
* 14360 14459: gap of unknown length
* 14460 15625: contig of 1166 bp in length
* 15626 15725: gap of unknown length
* 15726 17134: contig of 1409 bp in length
* 17135 17234: gap of unknown length
* 17235 18728: contig of 1494 bp in length
* 18729 18828: gap of unknown length
* 18829 20366: contig of 1538 bp in length
* 20367 20466: gap of unknown length
* 20467 21995: contig of 1529 bp in length
* 21996 22095: gap of unknown length
* 22096 23292: contig of 1197 bp in length
* 23293 23392: gap of unknown length
* 23393 24954: contig of 1562 bp in length
* 24955 25054: gap of unknown length
* 25055 26072: contig of 1018 bp in length
* 26073 26172: gap of unknown length
* 26173 27383: contig of 1211 bp in length
* 27384 27483: gap of unknown length
* 27484 28663: contig of 1180 bp in length
* 28664 28763: gap of unknown length
* 28764 29866: contig of 1103 bp in length
* 29867 31660: gap of unknown length
* 31661 31760: gap of unknown length
* 31761 32761: contig of 1001 bp in length
* 32762 32861: gap of unknown length
* 32862 34306: contig of 1445 bp in length
* 34307 34406: gap of unknown length
* 34407 36015: contig of 1609 bp in length
* 36016 36115: gap of unknown length
* 36116 37733: contig of 1618 bp in length
* 37734 37833: gap of unknown length
* 37834 39472: contig of 1639 bp in length
* 39473 39572: gap of unknown length
* 39573 41586: contig of 2014 bp in length
* 41587 41686: gap of unknown length
* 41687 43005: contig of 1319 bp in length
* 43006 43105: gap of unknown length
* 43106 44919: contig of 1814 bp in length
* 44920 45019: gap of unknown length
* 45020 46210: contig of 1191 bp in length
* 46211 46310: gap of unknown length
* 46311 48629: contig of 2319 bp in length
* 48630 48729: gap of unknown length
* 48730 50329: contig of 1600 bp in length
* 50330 50429: gap of unknown length
* 50430 51800: contig of 1371 bp in length
* 51801 51900: gap of unknown length
* 51901 54894: contig of 2994 bp in length
* 54895 54994: gap of unknown length
* 54995 57436: contig of 2442 bp in length
* 57437 57536: gap of unknown length
* 57537 59281: contig of 1745 bp in length
* 59282 59381: gap of unknown length
* 59382 61694: contig of 2313 bp in length
* 61695 61794: gap of unknown length
* 61795 64543: contig of 2749 bp in length
* 64544 64643: gap of unknown length

```

```

* 64644 66047: contig of 1404 bp in length
* 66048 66147: gap of unknown length
* 66148 68418: contig of 2271 bp in length
* 68419 68518: gap of unknown length
* 68519 70306: contig of 1788 bp in length
* 70307 72406: gap of unknown length
* 72407 72336: contig of 1930 bp in length
* 72337 72436: gap of unknown length
* 72437 74495: contig of 2059 bp in length
* 74496 74595: gap of unknown length
* 74596 76864: contig of 2269 bp in length
* 76865 76964: gap of unknown length
* 76965 78379: contig of 1415 bp in length
* 78380 78480: gap of unknown length
* 78480 80190: contig of 1711 bp in length
* 80191 80290: gap of unknown length
* 80291 83933: contig of 3643 bp in length
* 83934 84033: gap of unknown length
* 84034 85931: contig of 1898 bp in length
* 85932 86031: gap of unknown length
* 86032 88087: contig of 2056 bp in length
* 88088 88187: gap of unknown length
* 88188 90268: contig of 2081 bp in length
* 90269 90368: gap of unknown length
* 90369 92824: contig of 2456 bp in length
* 92825 92825: gap of unknown length
* 92925 95808: contig of 2884 bp in length
* 95809 95908: gap of unknown length
* 95909 98559: contig of 2651 bp in length

Query Match
Best Local Similarity 50.0%; Score 43.2; DB 2; Length 184700;
Matches 108; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 426 TCTTGTGATGGTGGAAATGGGACATCTCTGCTTAAACAGAAACATTCTTGTTC 485
DB 81232 TCCCTACTAGCTGATTTAGATGCTTACCTCAAGCCCTGTTCTGGACTTGT 81291
QY 486 GAGTAGACATCTGCTCTCTCTAGAGTAAGTTACCTCAGTCTCTGTCGT 545
DB 81292 CAGACATCTGATCTGTTGCTCTGCTGGAGGAGGAGGATCCCAATCTCT 81351
QY 546 GAGATTCTCTCTCTCTCGAGACCATCTGCTTAACTGAGGAGGAGTTC 605
DB 81352 GCCCTTTGCTGCTGCTGCTGAGGACCATCTCTATAGTCTATATGAGTATC 81411
QY 606 ACAGGTTCTCTCTCTTATCTTTGTTT 641
DB 81412 ACAGTCTCCACATAGTTTGTGTTT 81447

RESULT 13
AL390074 138006 bp DNA linear PRI 05-APR-2001
LOCUS Human DNA sequence from clone RP11-294H11 on chromosome 6, complete
DEFINITION sequence.
ACCESSION AL390074
VERSION AL390074.17 GI:13560018
KEYWORDS HTG.
SOURCE Human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 138006)
REFERENCE Tracey, A.
Direct Submission
Submitted (05-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquer@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Apr 6, 2001 this sequence version replaced GI:13396608.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the

```



Query Match 10.7% Score 71.6; DB 6; Length 2887;  
 Best Local Similarity 94.9%; Pred. No. 2.9e-11;  
 Matches 74; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 573 GATTCGCTTACGCTGGAGGAGAGTGTTCACAGGTTCTGCTCTTTATCTTTGT 632  
 1 GATTCGCTTCTGCTGGAGGAGAGTGTTCACAGGTTCTGCTCTTTATCTTTGT 60

QY 633 GTTTTTCGAGCCATG 650  
 61 GTTTTTCAGCCCTG 78

Db 61 GTTTTTCAGCCCTG 78

RESULT 9  
 AX134419 2887 bp DNA linear PAT 29-MAY-2001  
 DEFINITION Sequence 15 from Patent WO0129232.  
 ACCESSION AX134419  
 VERSION AX134419.1 GI:14271065  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 2887)  
 AUTHORS Cordell, B. and Li, Y.  
 TITLE Functional cloning of genes encoding proteins/enzymes involved in  
 proteolytic cleavage  
 JOURNAL Patent: WO 0129232-A 15 26-APR-2001;  
 Scios Inc. (US)

FEATURES  
 source Location/Qualifiers  
 1..2887  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 292..1731  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAC39526.1"  
 /db\_xref="GI:14271066"  
 /translation="MDPSRLVYDIGEOLDSEDLASLFLSLDYIPORKEPIKDALML  
 FORLOKRMEESENLSFKELLPRIINRLDILITVITRKEMERELOTPGSAOISAV  
 VALYQISEVSRSELSRFLQELISCKLDDNNLIDITENKRYILDEGALDIL  
 KRVCQINKSLKTIINDYEFSEKSSSLESPDEFSGEELCGVMTISDPRQDSE  
 SOTLDKRVQMSKPRGYCLIIINNFARAKRPKLHSIRDRNGHLDGALTTTFEE  
 LHFELKPHDQCTVEQIYEILKIYQLMDHSMDCFTCCILSHDGKGIYGTGQEARLY  
 ELTSQFTGLKCPISLAGKPKVFFIACQCGDNVOKGIPEVTDSEOPYLEMDLSPQTRY  
 IPDEADFLIGMATVNNCVSYRNPAEGTYIOSLCQLRRCRPGDDILITILEVNEYV  
 SNKDKKNNKGMPOPTFLRKILVFPD"

BASE COUNT 853 a 595 c 625 g 814 t

ORIGIN  
 Query Match 10.7% Score 71.6; DB 6; Length 2887;  
 Best Local Similarity 94.9%; Pred. No. 2.9e-11;  
 Matches 74; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 573 GATTCGCTTACGCTGGAGGAGAGTGTTCACAGGTTCTGCTCTTTATCTTTGT 632  
 1 GATTCGCTTCTGCTGGAGGAGAGTGTTCACAGGTTCTGCTCTTTATCTTTGT 60

Db 1 GATTCGCTTCTGCTGGAGGAGAGTGTTCACAGGTTCTGCTCTTTATCTTTGT 60

QY 633 GTTTTTCGAGCCATG 650  
 61 GTTTTTCAGCCCTG 78

Db 61 GTTTTTCAGCCCTG 78

RESULT 10  
 HSMACHAL 2887 bp mRNA linear PRI 25-SEP-1997  
 LOCUS HSMACHAL  
 DEFINITION H. sapiens mRNA for MACH-alpha-1 protein.  
 ACCESSION X88172  
 VERSION X88172.1 GI:1403318  
 KEYWORDS cell death; MACH-alpha-1; protease.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 2887)  
 AUTHORS Boidin, M.P., Goncharov, T.M., Golse, Y.V. and Wallach, D.  
 TITLE Involvement of MACH, a novel MOR1/RADP-interacting protease, in  
 Fas/Apo-1- and TNF receptor-induced cell death  
 JOURNAL Cell 85 (6), 803-815 (1996)  
 MEDLINE 96279826  
 PUBMED 8681376

REFERENCE 2 (bases 1 to 2887)  
 AUTHORS Wallach, D.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-MAY-1996) D. Wallach, The Weizmann Institute, Dept of  
 Membrane Research & Biophysics, Rehovot 76100, ISRAEL

REFERENCE 3 (bases 1 to 2887)  
 AUTHORS Wallach, D.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-JUN-1996) D. Wallach, The Weizmann Institute, Dept of  
 Membrane Research & Biophysics, Rehovot 76100, ISRAEL

REFERENCE 4 (bases 1 to 2887)  
 AUTHORS Mandruzzato, S., Brasseur, F., Andry, G., Boon, T. and van der  
 Bruggen, P.  
 TITLE A CASP-8 mutation recognized by cytolytic T lymphocytes on a human  
 head and neck carcinoma  
 JOURNAL J. Exp. Med. 186 (5), 785-793 (1997)  
 MEDLINE 97419196  
 PUBMED 9271594

FEATURES  
 source Location/Qualifiers  
 1..2887  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /tissue\_type="thymus"  
 292..1731  
 /function="protease"  
 /note="triggers cell death"  
 /codon\_start=1  
 /product="MACH-alpha-1"  
 /protein\_id="CA66853.1"  
 /db\_xref="GI:2440071"  
 /translation="MDPSRLVYDIGEOLDSEDLASLFLSLDYIPORKEPIKDALML  
 FORLOKRMEESENLSFKELLPRIINRLDILITVITRKEMERELOTPGSAOISAV  
 VALYQISEVSRSELSRFLQELISCKLDDNNLIDITENKRYILDEGALDIL  
 KRVCQINKSLKTIINDYEFSEKSSSLESPDEFSGEELCGVMTISDPRQDSE  
 SOTLDKRVQMSKPRGYCLIIINNFARAKRPKLHSIRDRNGHLDGALTTTFEE  
 LHFELKPHDQCTVEQIYEILKIYQLMDHSMDCFTCCILSHDGKGIYGTGQEARLY  
 ELTSQFTGLKCPISLAGKPKVFFIACQCGDNVOKGIPEVTDSEOPYLEMDLSPQTRY  
 IPDEADFLIGMATVNNCVSYRNPAEGTYIOSLCQLRRCRPGDDILITILEVNEYV  
 SNKDKKNNKGMPOPTFLRKILVFPD"

BASE COUNT 853 a 595 c 625 g 814 t

ORIGIN  
 Query Match 10.7% Score 71.6; DB 9; Length 2887;  
 Best Local Similarity 94.9%; Pred. No. 2.9e-11;  
 Matches 74; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 573 GATTCGCTTACGCTGGAGGAGAGTGTTCACAGGTTCTGCTCTTTATCTTTGT 632  
 1 GATTCGCTTCTGCTGGAGGAGAGTGTTCACAGGTTCTGCTCTTTATCTTTGT 60

Db 1 GATTCGCTTCTGCTGGAGGAGAGTGTTCACAGGTTCTGCTCTTTATCTTTGT 60

QY 633 GTTTTTCGAGCCATG 650  
 61 GTTTTTCAGCCCTG 78

Db 61 GTTTTTCAGCCCTG 78

RESULT 11  
 I66494 7218 bp DNA linear PAT 26-DEC-1997  
 LOCUS I66494  
 DEFINITION Sequence 14 from patent US 5670367.  
 ACCESSION I66494  
 VERSION I66494.1 GI:2724471  
 KEYWORDS

OY 561 TCTCGAGAGACAGATCTGCTTACGCTGGAGGAGAGTGTTCACAGTTCTCCTC 618  
 Db 61 TCTCGAGAGACAGATCTGCTTACGCTGGAGGAGAGTGTTCACAGATTATATTC 118

RESULT 6  
 AF422929 777 bp mRNA linear PRI 03-APR-2002  
 LOCUS Homo sapiens clone 4 procaspase-8 (CASP8) mRNA, partial cds;  
 DEFINITION alternatively spliced.  
 ACCESSION AF422929  
 VERSION AF422929.1 GI:19401529  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
 1 (bases 1 to 777)  
 Breckenridge,D.G., Nguyen,M., Kuppig,S., Reth,M. and Shore,G.C.  
 The procaspase-8 isoform, procaspase-8L, recruited to the BAP31  
 complex at the endoplasmic reticulum  
 Proc. Natl. Acad. Sci. U.S.A. 99 (77), 4331-4336 (2002)

JOURNAL  
 MEDLINE 21927603  
 PUBMED 11917123

REFERENCE  
 2 (bases 1 to 777)  
 Breckenridge,D.G., Nguyen,M., Kuppig,S., Reth,M. and Shore,G.C.  
 Direct Submission  
 Submitted (24-SEP-2001) Biochemistry, McGill University, 3655  
 Promenade Sir William Osler, Montreal, PQ H3G1Y6, Canada

JOURNAL  
 TITLE Submitted Sir William Osler, Montreal, PQ H3G1Y6, Canada  
 LOCATION/Qualifiers

FEATURES  
 source  
 1..777  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="4"  
 1..>777  
 /gene="CASP8"  
 266..>777  
 /gene="CASP8"  
 /note="caspase precursor; alternatively spliced"  
 /codon\_start=1  
 /product="procaspase-8"  
 /protein\_id="AAU87632.1"  
 /db\_xref="GI:19401530"  
 /translation="MDFSRNLYDIGEOLDSSEDLASIKFLSLDIYIPQKOEPIKDALM  
 FORLOEKRMLEESNLSFLKELFRINRDLITLNTREEMERELQTFGRQISAYR  
 VMLYQISEEVSRELSRFSKFLLOEISKCKLDDNMLDIFTEMERVILGSKLDIL  
 KRYCAQINKSL"

BASE COUNT 218 a 156 c 203 g 194 t

ORIGIN

Query Match 10.7%; Score 71.6; DB 9; Length 777;  
 Best Local Similarity 94.9%; Pred. No. 2.5e-11;  
 Matches 74; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 573 GATTGCGCTTACGCTGGAGGAGAGTGTTCACAGTTCCTCTTATCTTTGT 632  
 Db 1 GATTGCGCTTCTGCTGGAGGAGAGTGTTCACAGTTCCTCTTATCTTTGT 60

OY 633 GTTTTTCGAGCCATG 650  
 Db 61 GTTTTTCGAGCCCTG 78

RESULT 7  
 AF422925 2503 bp mRNA linear PRI 03-APR-2002  
 LOCUS Homo sapiens procaspase-8L (CASP8) mRNA, complete cds;  
 DEFINITION alternatively spliced.  
 ACCESSION AF422925  
 VERSION AF422925.1 GI:19401518  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.

REFERENCE  
 1 (bases 1 to 2503)  
 Breckenridge,D.G., Nguyen,M., Kuppig,S., Reth,M. and Shore,G.C.  
 The procaspase-8 isoform, procaspase-8L, recruited to the BAP31  
 complex at the endoplasmic reticulum  
 Proc. Natl. Acad. Sci. U.S.A. 99 (77), 4331-4336 (2002)

JOURNAL  
 MEDLINE 21927603  
 PUBMED 11917123

REFERENCE  
 2 (bases 1 to 2503)  
 Breckenridge,D.G., Nguyen,M., Kuppig,S., Reth,M. and Shore,G.C.  
 Direct Submission  
 Submitted (24-SEP-2001) Biochemistry, McGill University, 3655  
 Promenade Sir William Osler, Montreal, PQ H3G1Y6, Canada

JOURNAL  
 TITLE Submitted Sir William Osler, Montreal, PQ H3G1Y6, Canada  
 LOCATION/Qualifiers

FEATURES  
 source  
 1..2503  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..2503  
 /gene="CASP8"  
 115..1731  
 /gene="CASP8"  
 /note="caspase precursor; alternatively spliced"  
 /codon\_start=1  
 /product="procaspase-8L"  
 /protein\_id="AAU87628.1"  
 /db\_xref="GI:19401519"  
 /translation="MEGGRARVYTESRNFGLAPTPPAEVEIGRLDSESTAY  
 FGKGDADILPFRKMDSRNLYDIGEOLDSSEDLASIKFLSLDIIPRKQEPIDALM  
 LFORLOEKRMLEESNLSFLKELFRINRDLITLNTREEMERELQTFGRQISAY  
 RMLYQISEEVSRELSRFSKFLLOEISKCKLDDNMLDIFTEMERVILGSKLDI  
 LKRYCAQINKSLKIINDYEEFSKERSLSLSDPESNGEELGVMTIDSPDEQDS  
 ESQTDKRYOKRSKRGYCLIIINHNPKAEKXPKLSIDRNGTGLDAGALTPE  
 ELHPIKRPDCTVQGLYEILKIYIOMDSMGDFICCIISHGKGLITGQGEARI  
 YELNSQFTGLKPSLAKGRKRYFTQACQGDGVQGIYEPDSEFQPIEMDSRPR  
 YIPDADFLGMATVNNCVSYRNPBEGTWIQLSCQSLREKCPRGDDILITLLEVNE  
 VSNKDKRNMGMKQMPQPTFLRKRLVPSD"

BASE COUNT 726 a 528 c 593 g 656 t

ORIGIN

Query Match 10.7%; Score 71.6; DB 9; Length 2503;  
 Best Local Similarity 94.9%; Pred. No. 2.8e-11;  
 Matches 74; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 573 GATTGCGCTTACGCTGGAGGAGAGTGTTCACAGTTCCTCTTATCTTTGT 632  
 Db 1 GATTGCGCTTCTGCTGGAGGAGAGTGTTCACAGTTCCTCTTATCTTTGT 60

OY 633 GTTTTTCGAGCCATG 650  
 Db 61 GTTTTTCGAGCCCTG 78

RESULT 8  
 AR211527 2887 bp DNA linear PAT 20-JUN-2002  
 LOCUS AR211527 Sequence 14 from patent US 6399327.  
 DEFINITION AR211527  
 ACCESSION AR211527  
 VERSION AR211527.1 GI:21514872  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 Unclassified.  
 1 (bases 1 to 2887)  
 Wallach,D., Bouldin,M., Goncharov,T. and Golsteyn,Y.V.  
 Modulators of the function of FAS receptors and other proteins  
 Patent: US 6399327-A 14 04-JUN-2002;  
 LOCATION/Qualifiers

FEATURES  
 source  
 1..2887  
 /organism="unknown"  
 /db\_xref="taxon:9606"  
 853 a 595 c 625 g 814 t

BASE COUNT 853 a 595 c 625 g 814 t

ORIGIN

QY		377	CAGGCGCTGGGGCTTAGTTTGTCACAGTCCATGAATTGTCTGCACATCCCTCTTCGAATG	436
Dd		121	CAGGCGCTGGGGCTTAGTTTGACAGTCCATGAATGTCTGCACATCCCTCTTCGAATG	180
QY		437	GTTGCAATTTGGCATCTCTGTCTTTAAACAGAGAACAATTTCTTGTCAGATGATCAT	496
Dd		181	TT	239
QY		497	CTCTGTTGCTTTAGAGAGTAAGTAAGTTAACCTCAGTTCCTCTCTGTGGTGAAGTTTCTC	556
Dd		240	CTCTGTTGCTTTAGAGAGTAAGTAAGTTAACCTCAGTTCCTCTCTGTGGTGAAGTTTCTC	299
QY		557	TTCCTCTGGAGACACAGATTCTGCCCTTTACGCTGAGAGGAAAGTCTTTACAGGTTCC	616
Dd		300	TTCTCTCTGGAGACACAGATTCTGCCCTTTACGCTGAGAGGAAAGTCTTTACAGGTTCTC	359
QY		617	TCCTTTAATCTTTGTGTTTCTTTTTCGAGCGATG	650
Dd		360	TCCTTTAATCTTTGTGTTTCTTTTTCGAGCGATG	393
RESULT 5				
AF422926		413 bp	mRNA.	linear PRI 03-APR-2002
LOCUS				
DEFINITION		Homo sapiens clone 1 procaspace-8 (CASP8) mRNA, partial cds;		
ACCESSION		AF422926		
VERSION		AF422926.1		
KEYWORDS		GI:19401520		
SOURCE				
ORGANISM		Homo sapiens.		
		Homo sapiens		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE				
AUTHORS		1 (bases 1 to 413)		
TITLE		Breckenridge,D.G., Nguyen,M., Kupplig,S., Reth,M. and Shore,G.C.		
JOURNAL		The procaspace-8 isoform, procaspace-8L, recruited to the Bap31		
MEDLINE		complex at the endoplasmic reticulum		
PUBMED		Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4331-4336 (2002)		
21927603				
11917123				
2 (bases 1 to 413)				
Breckenridge,D.G., Nguyen,M., Kupplig,S., Reth,M. and Shore,G.C.				
Direct Submission				
Submitted (24-Sep-2001) Biochemistry, McGill University, 3655				
Promenade Sir William Osler, Montreal, PQ H3G1Y6, Canada				
Location/Qualifiers				
1..413				
/organism="Homo sapiens"				
/db_xref="taxon:9606"				
/clone="1"				
1..>413				
/gene="CASP8"				
136..>413				
/gene="CASP8"				
/note="caspase precursor; alternatively spliced"				
/codon_start=1				
/product="procaspace-8"				
/protein_id="ABL87629.1"				
/db_xref="GI:19401521"				
/translation="MDFSRNLYDIGEQLDSEDLASLKFLSYLPORKOEPFKDALMLN				
FRLQEKRMLESNIISLFKELLFPINRDLLITLYLNTRKEMERELDTP"				
BASE COUNT		115 a	87 c	96 g
ORIGIN				
Query Match		16.2%	Score 108.4:	DB 9: Length 413:
Best Local Similarity		94.9%	Pred. No. 3.9e-23;	
Matches 112; Conservative		0;	Mismatches 6;	Indels 0; Gaps 0
QY		501	GTTCTGCTTTAGAGAGTAAGTAAGTTAACCTCAGTTCCTCTGTGGTGAAGTTTCTCTTC	560
Dd		1	GTTCTGCTTTAGAGAGTAAGTAAGTTAACCTCAGTTCCTCTGTGGTGAAGTTTCTCTTC	60

misc_feature	4128. .4340	/note="match to EST AV760501 (NID:g10918349)"
misc_feature	4184. .4742	

```

misc_feature /note="match to EST AI052220 (NID:g3308241) o21c07.x1"
4339 .4903
/note="similar to Homo sapiens EST AU117593
(NID:g10932573)"
misc_feature 4485 .4491
/note="similar to Homo sapiens EST BE788113
(NID:g10209311)"
misc_feature 4657 .5468
/note="similar to Homo sapiens EST AL563032
(NID:g12612044)"
misc_feature 4712 .5400
/note="match to EST AU131936 (NID:g10992290)"
7470 .5582
/note="similar to Homo sapiens EST BG164297
(NID:g12671000)"
misc_feature 4963 .5682
/note="similar to Homo sapiens EST AU124988
(NID:g10949704)"
misc_feature 5106 .5533
/note="similar to Mus musculus EST A116325 (NID:g3516649)
ue97c04.y1"
misc_feature 5169 .5675
/note="similar to Homo sapiens EST AU144381
(NID:g11005502)"
misc_feature 5194 .5761
/note="similar to Homo sapiens EST BF529656
(NID:g11617019)"
misc_feature 5237 .6019
/note="similar to Homo sapiens EST AU130508
(NID:g10990862)"
misc_feature 5238 .6043
/note="similar to Mus musculus EST BF166846
(NID:g11047198)"
misc_feature 5326 .5653
/note="similar to Homo sapiens EST BG390376
(NID:g13283824)"
misc_feature 5361 .6051
/note="similar to Homo sapiens EST AU155897
(NID:g10896436)"
misc_feature 5361 .5812
/note="similar to EST BG862638 (NID:g14213176)"
5420 .6048
/note="similar to Homo sapiens EST BE788113
(NID:g10209311)"
misc_feature 5551 .6327
/note="similar to Homo sapiens EST AU130771
(NID:g10991125)"
misc_feature 5615 .5584
/note="similar to Homo sapiens EST AA280659 (NID:g1923454)

```

	Query Match	77.0%	Score 516;	DB 9:	Length 181150;	
	Best Local Similarity	94.4%	Pred. No. 2.9e-153;			
	Matches 566; Conservative	0;	Mismatches 30;	Indels 4;	Gaps 3;	
QY	49	AATTTAAGAACACGCGCCAGAGTGGGAACCTCAGCCTTAGCACGCGGTTCATCCGAGACAG	108			
Db	159872	AAACTGAGACACTGGCCCAAGATGSGAATCTAGCCTCAGACGGCGTGAGCC - GAGCAG	159815			
QY	109	GCGTAAGCCCAAGTACGAATGAAACCCAGACCACCTCTCTCTTTTTTTCTGAAAGATCTACCC	168			
Db	159814	GCGTAAGCCCAAGTACGAGCGCTGACGCCAGAACACTCTCTCTTTTT - TCTGAACAATCTACCT	159756			
QY	169	GCATTTCACCCACACGAGGCGTCACTTTACCCAGTCCGCGCGAGGAGAGAGAGAGCGTGTCT	228			
Db	159755	ACATTTCAGCTACAGGGCTGGGCTTTTACCACAGTCCGGCGGAGGAGGAGAGCGTGTCT	159656			
QY	229	GTCACCTTCAGCTGTGAGGCTTGTATCATCAGAAGCAAAGGAACTTCTATTCCAGACCCCTTT	268			
Db	159695	GTCACCTTCAGTGTGTGAGGCTTGTATCATCAGAAGCAAAGGAACTCTCTATTCCAGACCCCTTT	159638			
QY	289	GCAAGAAAGATGAGCATATTACTTCGCGCGCAGAGGGTPTTATTTACTATAATGAGTCA	348			
Db	159635	GCAAGAAAGATGAGCATATTACTTCGCGCGCAGAGGGTPTTATTTACTATAATGAGTCA	159576			



FEATURES		Location/Qualifiers
source		1. 179216 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="2" /clone="RP11-155P18"
misc_feature		1. 82838 /note="assembly_name:Contig10 clone_end:87 vector_side:left"
misc_feature		82939..179216 /note="assembly_name:Contig11 clone_end:SP6 vector_side:left"
BASE COUNT	47524 a 37637 c 38145 g 55760 t	150 others
ORIGIN		
Query Match	77.0% ; Score 516; DB 2; Length 179216;	
Best Local Similarity	94.4% ; Pred. No. 2.9e-153;	
Matches 568; Conservative	0; Mismatches 30; Indels 4; Gaps 3;	
QY	49	AAATTAGACAGGGCCAAAGATGGGAACACAGCCTAGACACGGCTGTGATCCGACAG 108
Db	152317	AAACTGGAAGACATGSCCAAGATGGGAATCAGCTTACAGAGGGCTGAGCC--GACAG 152374
QY	109	GGCCTAAGCCAGTAGCAATGAACACAGACCCTCTCCCTTTTCTGGAAGATCAACC 168
Db	152375	GGCTTAAGCCAGTAGAGGGCTGAGCCAGCAACACTTCCTCTTTT--TGTGAACAATCACT 152433
QY	169	GCATTTCACCCACAGGGGCTGACTTTACCCAGTCGCGCGGGAGGAGAGAGGCTGGCT 228
Db	152434	ACATTTCACCTACAGGGGCTGACTTTACCCAGTCGCGCGGGAGGAGAGAGGCTGGCT 152493
QY	229	GTGACTTCAGTGGTGAAGTTTGATCAAGGCAAGGGAAGTTCCTATTCCAGACCCCTT 288
Db	152494	GTGACTTCAGTGGTGAAGTTTGATCAAGGCAAGGGAAGTTCCTATTCCAGACCCCTT 152553
QY	289	GCAAGAAAGAAATGCAATTAATCTTCCCGCCGACAGGGGTTATTATCTAAATGGAGTCA 348
Db	152554	GCAAGAAAGAAATGCAATTAATCTTCCCGCCGACAGGGGTTATTATCTAAATGGAGTCA 152613
QY	349	GTAATAATGTTTCCCAATAAAGCAGTCGAGGGCTCGGGCTTATGTTGACGTCATGA 408
Db	152614	GTAATAATGTTTCCCAATAAAGCAGTCGAGGGCTCGGGCTTATGTTGACGTCATGA 152673
QY	409	ATTGTCTGCACATCCCTCTTCTGATGAGTTGGGATCTGTTCCCTTTAAACA 468
Db	152674	ATTGTCTGCACATCCCTCTTCTGATGAGTTGGGATCTGTTCCCTTTAAACA 152732
QY	469	GGAACACATTTCTGTCGAGTGAGTCACTCGTCTCTCTTAAAGAAAGTTTACCCT 528
Db	152733	GGAACACATTTCTGTCGAGTGAGTCACTCGTCTCTCTTAAAGAAAGTTTACCCT 152792
QY	529	GCAAGTCTCTCTGTGTGTGAAGTTTCTCTTCTCTCGAGACCAAGATTCGCTTTACGC 588
Db	152793	GCAAGTCTCTCTGTGTGTGAAGTTTCTCTTCTCTCGAGACCAAGATTCGCTTTACGC 152852
QY	589	TGAGAGGAAGTGTGTTTCAAGAGTTCTCCCTTTTATCTTTTGTGTTTTTTTGAACCA 648
Db	152853	TGAGAGGAAGTGTGTTTCAAGAGTTCTCCCTTTTATCTTTTGTGTTTTTTTGAACCC 152912
QY	649	TT 650
Db	152913	TG 152914
RESULT 3		
AC007256/c	AC007256	181150 bp DNA linear PRI 07-NOV-2001
LOCUS	DEFINITION	Homo sapiens BAC clone RP11-575C6 from 2, complete sequence.
AC007256	AC007256	
ACCESSION	AC007256.5	GI:14527821
VERSION		
KEYWORDS	HTG.	

```

SOURCE      Homo sapiens.
ORGANISM    Homo sapiens.
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE   1 (bases 1 to 181150)
AUTHORS     Sulston,J.E. and Waterston,R.
TITLE       Toward a complete human genome sequence
JOURNAL     Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE     99063792
PUBMED      9847074

REFERENCE   2 (bases 1 to 181150)
AUTHORS     Du,F., Laplant,Y., Doeber,A. and Moore,B.
TITLE       The sequence of Homo sapiens BAC clone RP11-575C6
JOURNAL     Unpublished
REFERENCE   3 (bases 1 to 181150)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (07-APR-1999) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis
             MO 63108, USA
REFERENCE   4 (bases 1 to 181150)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (19-APR-2001) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
REFERENCE   5 (bases 1 to 181150)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (20-APR-2001) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
REFERENCE   6 (bases 1 to 181150)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (07-JUN-2001) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
REFERENCE   7 (bases 1 to 181150)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (03-JUL-2001) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
REFERENCE   8 (bases 1 to 181150)
AUTHORS     Waterston,R.
TITLE       Direct Submission
JOURNAL     Submitted (07-NOV-2001) Department of Genetics, Washington
             University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
             On Jun 7, 2001 this sequence version replaced gi:13677176.

COMMENT     ----- Genome Center
             Center: Washington University Genome Sequencing Center
             Center code: WUGSC
             Web site: http://genome.wustl.edu/gsc
             Contact: saplens@watsn.wustl.edu
             ----- Summary Statistics
             Center project name: H_NH0575C06
             -----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:21:06 ; Search time 1995 Seconds  
(without alignments)  
9773.875 Million cell updates/sec

Title: US-09-477-082-1

Perfect score: 670  
Sequence: 1 aaagcctcacaagacgatt.....ggggtaataaagcgttt 670

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vl.\*  
30: em\_hg\_hum.\*  
31: em\_hg\_inv.\*  
32: em\_hg\_other.\*  
33: em\_hg\_mus.\*  
34: em\_hg\_pln.\*  
35: em\_hg\_rnd.\*  
36: em\_hg\_mam.\*  
37: em\_hg\_vrt.\*  
38: em\_sy.\*  
39: em\_hgo\_hum.\*  
40: em\_hgo\_mus.\*  
41: em\_hgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	516	77.0 6502	9 AB03898052	AB038981 Homo sapi
2	516	77.0 179216	2 AC074016	AC074016 Homo sapi
3	516	77.0 181150	9 AC0007256	AC0007256 Homo sapi
4	366	54.6 982	9 AF210257	AF210257 Homo sapi
5	108.4	16.2 413	9 AF422926	AF422926 Homo sapi
6	71.6	10.7 777	9 AF422929	AF422929 Homo sapi
7	71.6	10.7 2503	9 AF422925	AF422925 Homo sapi
8	71.6	10.7 2887	6 AR211527	AR211527 Sequence
9	71.6	10.7 2887	6 AX134419	AX134419 Sequence
10	71.6	10.7 2887	6 HSMACHA1	X98172 H.sapiens m
11	49.6	7.4 7238	6 T66494	T66494 Sequence 14
12	43.2	6.4 184700	2 AC112822	AC112822 Rattus no
13	42.8	6.4 138006	2 AL330074	AL330074 Human DNA
14	42.8	6.4 223387	2 AC079470	AC079470 Mus muscu
15	42.2	6.3 186882	9 AL136307	AL136307 Human DNA
16	41.8	6.2 148334	2 AC087106	AC087106 Homo sapi
17	40.6	6.1 49430	2 AC100434	AC100434 Mus muscu
18	40	6.0 121538	9 AC128707	AC128707 Homo sapi
19	40	6.0 145002	2 AC018372	AC018372 Homo sapi
20	40	6.0 148208	2 AC010951	AC010951 Homo sapi
21	40	6.0 160325	2 AC073044	AC073044 Homo sapi
22	40	6.0 176887	2 AC041032	AC041032 Homo sapi
23	40	5.8 186581	2 AL1506810	AL1506810 Homo sapi
24	38.8	5.8 209518	2 AC098348	AC098348 Rattus no
25	38.6	5.8 36037	9 AC003966	AC003966 Homo sapi
26	38.6	5.8 157443	9 AC109352	AC109352 Homo sapi
27	38.6	5.8 158804	2 AC123019	AC123019 Rattus no
28	38.4	5.7 141400	9 AP004581	AP004581 Homo sapi
29	38.2	5.7 137227	2 HUMH97G12	HUMH97G12 Homo sapi
30	38.2	5.7 69305	9 AL357496	AL357496 Human DNA
31	38.2	5.7 72045	9 AC092214	AC092214 Homo sapi
32	38.2	5.7 164018	2 AC074242	AC074242 Homo sapi
33	38.2	5.7 172387	2 AC104753	AC104753 Mus muscu
34	38.2	5.7 209685	2 AC026231	AC026231 Mus muscu
35	38.2	5.7 114409	2 AL161742	AL161742 Human DNA
36	38	5.7 137227	2 AL356914	AL356914 Homo sapi
37	38	5.7 158622	2 AC117302	AC117302 Rattus no
38	38	5.7 160175	2 AC090178	AC090178 Homo sapi
39	38	5.7 163502	9 AC023456	AC023456 Homo sapi
40	38	5.7 173386	9 AC016343	AC016343 Homo sapi
41	38	5.7 182107	9 AC015798	AC015798 Homo sapi
42	38	5.6 110042	9 AL590788	AL590788 Human DNA
43	37.8	5.6 146098	9 AC074112	AC074112 Homo sapi
44	37.8	5.6 182431	2 AC068380	AC068380 Homo sapi
45	37.8	5.6 182431	2 AC068380	AC068380 Homo sapi

## ALIGNMENTS

RESULT 1  
AB03898052  
LOCUS AB03898052 6502 bp DNA linear PRI 23-FEB-2001  
DEFINITION Homo sapiens gene for caspase-8, exon 3, exon 4.  
ACCESSION AB038981  
VERSION  
KEYWORDS  
SEGMENT  
SOURCE  
ORGANISM  
Homo sapiens DNA.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1 Hadano, S., Yanagisawa, Y., Skaug, J., Fichter, K., Nasir, J.,  
Martindale, D., Koop, B.F., Scherer, S.W., Nicholson, D.W.,

Search completed: July 8, 2003, 06:04:42  
Job time : 25.6593 secs

```

: Sequence 868, Application US/09969373
: Patent No. US2002013852A1
: GENERAL INFORMATION:
: APPLICANT: Efeitz, Roger J.
: APPLICANT: Hauge, Brian M.
: TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
: FILE REFERENCE: 38-10(52679)/A
: CURRENT APPLICATION NUMBER: US/09/969,373
: CURRENT FILING DATE: 2001-10-02
: PRIOR APPLICATION NUMBER: US 09/754,853
: PRIOR FILING DATE: 2001-01-05
: PRIOR APPLICATION NUMBER: US 09/760,427
: PRIOR FILING DATE: 2001-01-13
: PRIOR APPLICATION NUMBER: US 09/855,768
: PRIOR FILING DATE: 2001-05-15
: NUMBER OF SEQ ID NOS: 4593
: SEQ ID NO 868
: LENGTH: 132
: TYPE: DNA
: ORGANISM: Glycine max
US-09-969-373-868

```

```

Query Match      72.8%; Score 18.2; DB 10; Length 132;
Best Local Similarity 87.0%; Pred. No. 6.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY      3 ATATATCTACATTCACAAACA 25
      ||||||| ||||| |||
Db      53 ATATATATATATATTCACAAATCA 31

```

```

RESULT 15
US-10-066-543-2507
: Sequence 2507, Application US/10066543
: Publication No. US20030087818A1
: GENERAL INFORMATION:
: APPLICANT: Jiang, Yugu
: APPLICANT: Pyle, Ruth A.
: APPLICANT: Xu, Jiangchun
: APPLICANT: Indrias, Carol Joseph
: APPLICANT: Lodes, Michael J.
: APPLICANT: Secrist, Heather
: APPLICANT: Carter, Darrick
: APPLICANT: Fanger, Gary R.
: APPLICANT: Smith, Carole L.
: APPLICANT: Durham, Margarita
: APPLICANT: Stolk, John A.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
: FILE REFERENCE: 210121.563
: CURRENT APPLICATION NUMBER: US/10/066,543
: CURRENT FILING DATE: 2002-01-31
: NUMBER OF SEQ ID NOS: 3417
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2507
: LENGTH: 234
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 115, 127, 156, 163, 208
: OTHER INFORMATION: n = A,T,C or G
US-10-066-543-2507

```

```

Query Match      72.8%; Score 18.2; DB 9; Length 234;
Best Local Similarity 87.0%; Pred. No. 7.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY      3 ATATATCTACATTCACAAACA 25
      ||||||| ||||| |||
Db      83 ATATATATATACATTGAAAAA 105

```

;; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
;; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
;; FILE REFERENCE: MRI-049  
;; CURRENT APPLICATION NUMBER: US/10/198,846  
;; CURRENT FILING DATE: 2002-07-18  
;; PRIOR APPLICATION NUMBER: 60/306,220  
;; PRIOR FILING DATE: 2001-07-18  
;; NUMBER OF SEQ ID NOS: 14084  
;; SOFTWARE: fastseq for windows version 4.0  
;; SEQ ID NO 13758  
;; LENGTH: 4020  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: 1, 2, 1631, 4018, 4019, 4020  
;; OTHER INFORMATION: n = A,T,C or G  
US-10-198-846-13758

Query Match 74.4%; Score 18.6; DB 9; Length 4020;  
Best Local Similarity 84.0%; Pred. No. 7.5e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATTCATAAACA 25  
|||||  
DB 1713 CCATATATATATATATCCAGACAA 1737

RESULT 11  
US-10-239-676-78/C  
;; Sequence 78, Application US/10239676  
;; Publication No. US20030082609A1  
;; GENERAL INFORMATION:  
;; APPLICANT: OLEK, Alexander  
;; APPLICANT: PIEPENBROCK, Christian  
;; APPLICANT: BERLIN, Kurt  
;; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
;; FILE REFERENCE: 5013.1003  
;; CURRENT APPLICATION NUMBER: US/10/239,676  
;; CURRENT FILING DATE: 2002-09-24  
;; PRIOR APPLICATION NUMBER: PCT/EP01/03968  
;; DE 10018058.8  
;; DE 10019173.8  
;; DE 10032529.7  
;; DE 10043826.1  
;; PRIOR FILING DATE: 2001-04-06  
;; 2000-04-06  
;; 2000-04-07  
;; 2000-06-30  
;; 2000-09-01  
;; NUMBER OF SEQ ID NOS: 228  
;; SEQ ID NO 78  
;; LENGTH: 6071  
;; TYPE: DNA  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
;; NAME/KEY: unsure  
;; LOCATION: (2618, 2780)  
US-10-239-676-78

Query Match 74.4%; Score 18.6; DB 9; Length 6071;  
Best Local Similarity 84.0%; Pred. No. 7.9e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATTCATAAACA 25  
|||||  
DB 3475 CAAATAATATTACATTCATAACAA 3451

RESULT 12

US-09-754-853A-2  
;; Sequence 2, Application US/09754853A  
;; Publication No. US20030005491A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Hauge, Brian M.  
;; APPLICANT: Parnell, Laurence D.  
;; APPLICANT: Parsons, Jeremy D.  
;; APPLICANT: Wang, Ming Li  
;; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
;; FILE REFERENCE: 38-10(15810)B  
;; CURRENT APPLICATION NUMBER: US/09/754,853A  
;; CURRENT FILING DATE: 2001-01-05  
;; PRIOR APPLICATION NUMBER: US 60/174,880  
;; PRIOR FILING DATE: 2000-01-07  
;; NUMBER OF SEQ ID NOS: 1119  
;; SEQ ID NO 2  
;; LENGTH: 335913  
;; TYPE: DNA  
;; ORGANISM: Glycine max  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (45163)..(45314)..(45450)..(45509)..(46941)..(48763)..(48975)..(49573)  
;; OTHER INFORMATION: Clone ID: 240017\_region\_G3  
US-09-754-853A-2

Query Match 74.4%; Score 18.6; DB 9; Length 335913;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATTCATAAACA 25  
|||||  
DB 20113 CGATATATATATACATAAACA 201137

RESULT 13  
US-09-754-853A-3  
;; Sequence 3, Application US/09754853A  
;; Publication No. US20030005491A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Hauge, Brian M.  
;; APPLICANT: Parnell, Laurence D.  
;; APPLICANT: Parsons, Jeremy D.  
;; APPLICANT: Wang, Ming Li  
;; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
;; FILE REFERENCE: 38-10(15810)B  
;; CURRENT APPLICATION NUMBER: US/09/754,853A  
;; CURRENT FILING DATE: 2001-01-05  
;; PRIOR APPLICATION NUMBER: US 60/174,880  
;; PRIOR FILING DATE: 2000-01-07  
;; NUMBER OF SEQ ID NOS: 1119  
;; SEQ ID NO 3  
;; LENGTH: 335913  
;; TYPE: DNA  
;; ORGANISM: Glycine max  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (46798)..(48763)..(48975)..(49573)  
;; OTHER INFORMATION: Clone ID: 240017\_region\_G3  
US-09-754-853A-3

Query Match 74.4%; Score 18.6; DB 9; Length 335913;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATTCATAAACA 25  
|||||  
DB 20113 CGATATATATATACATAAACA 201137

RESULT 14  
US-09-969-373-868/C

US-09-918-995-9724

Query Match 74.4% Score 18.6; DB 9; Length 469;  
Best Local Similarity 84.0%; Pred. No. 5.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATCAAAACAA 25  
130 CCATATATATATATATCCAGACAA 154

RESULT 7

US-09-918-995-13475  
Sequence 13475, Application US/09918995  
Publication No. US20030073623A1

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.

TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

FILE REFERENCE: 20411-756 FROM VARIOUS CDNA LIBRARIES

CURRENT APPLICATION NUMBER: US/09/918,995

PRIOR FILING DATE: 2001-07-30

PRIOR APPLICATION NUMBER: US/09/235,076

PRIOR FILING DATE: 1999-01-20

NUMBER OF SEQ ID NOS: 38054

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 13475

LENGTH: 487

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (1)...(487)

OTHER INFORMATION: n = A,T,C or G

US-09-918-995-13475

Query Match 74.4% Score 18.6; DB 9; Length 487;  
Best Local Similarity 84.0%; Pred. No. 5.7e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATCAAAACAA 25  
355 CCATATATATATATATCCAGACAA 379

RESULT 8

US-09-745-763-12  
Sequence 12, Application US/09745763  
Patent No. US20020065394A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

McCoy, John M.

LaVallie, Edward R.

Collins-Racie, Lisa A.

Evans, Cheryl

Merberg, David

Trecay, Maurice

Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

ENCODING THEM

NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS:

ADDRESS: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/745,763

FILING DATE: 18-Jun-2000

CLASSIFICATION: &lt;unknown&gt;

ATTORNEY/AGENT INFORMATION:

NAME: Sprunger, Suzanne A.

REGISTRATION NUMBER: 41,323

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8284

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 1463 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 12:

US-09-745-763-12

Query Match 74.4% Score 18.6; DB 10; Length 1463;

Best Local Similarity 84.0%; Pred. No. 6.6e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATCAAAACAA 25  
1167 CCATATATATATATATCCAGACAA 1191

RESULT 9

US-10-044-205A-21  
Sequence 21, Application US/10044205A  
Patent No. US20020123464A1

GENERAL INFORMATION:

APPLICANT: KAPPELLER-LIBERMANN, Rosana

BANDARU, Rajasekhar

TITLE OF INVENTION: 69087, 15821, and 15418, Methods and Compositions of Human Pro

FILE REFERENCE: 10147-5201

CURRENT APPLICATION NUMBER: US/10/044,205A

CURRENT FILING DATE: 2002-04-19

PRIOR APPLICATION NUMBER: US 60/242,428

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: US 60/241,884

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: US 60/241,877

NUMBER OF SEQ ID NOS: 44

SOFTWARE: PatentIn version 3.1

SEQ ID NO 21

LENGTH: 3003

TYPE: DNA

ORGANISM: Homo sapiens

US-10-044-205A-21

Query Match 74.4% Score 18.6; DB 12; Length 3003;

Best Local Similarity 84.0%; Pred. No. 7.2e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATCAAAACAA 25  
1997 CCATATATATATATATCCAGACAA 2021

RESULT 10

US-10-198-846-13758  
Sequence 13758, Application US/10198846  
Publication No. US20030099974A1

GENERAL INFORMATION:

APPLICANT: Lillie, James

Xu, Yongyao

Wang, Youzhen

Steinmann, Kathleen

```
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US 09/754,853
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 09/760,427
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US 09/855,768
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 877
; LENGTH: 200
; TYPE: DNA
; ORGANISM: Glycine max
; US-09-969-373-877

Query Match          74.4%  Score 18.6;  DB 10;  Length 200;
Best Local Similarity 84.0%  Pred. No. 5.1e+02;
Matches 21;  Conservative 0;  Mismatches 4;  Indels 0;  Gaps 0;

QY      1  CCATATATATCTACATTCAAACAA 25
        ||| ||| ||| ||| ||| ||| |||
Db      115 CGATATATATATACATTAACAA 139

RESULT 3
US-09-754-853A-149
; Sequence 149, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Soybean Cyst Nematode Resistance
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 149
; LENGTH: 221
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 240017_region_g3_201115_11
; US-09-754-853A-149

Query Match          74.4%  Score 18.6;  DB 9;  Length 221;
Best Local Similarity 84.0%  Pred. No. 5.2e+02;
Matches 21;  Conservative 0;  Mismatches 4;  Indels 0;  Gaps 0;

QY      1  CCATATATATCTACATTCAAACAA 25
        ||| ||| ||| ||| ||| ||| |||
Db      120 CGATATATATATACATTAACAA 144

RESULT 4
US-09-960-352-374/C
; Sequence 374, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengping
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 374
```

```
; LENGTH: 355
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 02-LIB34-005-Q1-E1-A9
; US-09-960-352-374

Query Match          74.4%  Score 18.6;  DB 10;  Length 355;
Best Local Similarity 84.0%  Pred. No. 5.5e+02;
Matches 21;  Conservative 0;  Mismatches 4;  Indels 0;  Gaps 0;

QY      1  CCATATATATCTACATTCAAACAA 25
        ||| ||| ||| ||| ||| ||| |||
Db      288 CCAATTAATCAACGTTCAAAACAA 264

RESULT 5
US-09-960-352-12299/C
; Sequence 12299, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengping
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12299
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (301)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 53-BOVMS1-006-Q1-E1-F2
; US-09-960-352-12299

Query Match          74.4%  Score 18.6;  DB 10;  Length 392;
Best Local Similarity 84.0%  Pred. No. 5.6e+02;
Matches 21;  Conservative 0;  Mismatches 4;  Indels 0;  Gaps 0;

QY      1  CCATATATATCTACATTCAAACAA 25
        ||| ||| ||| ||| ||| ||| |||
Db      249 CCAATTAATCAACGTTCAAAACAA 225

RESULT 6
US-09-918-995-9724
; Sequence 9724, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hysed, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: fastSeq for Windows Version 3.0
; SEQ ID NO 9724
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; LOCATION: (1)...(469)
; OTHER INFORMATION: n = A,T,C or G
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:57:09 ; Search time 12.6593 Seconds  
(Without alignments)  
3082.690 Million cell updates/sec

Title: US-09-477-082-32

Perfect score: 25

Sequence: 1 ccatatatactacatcaaaaca 25

Scoring table: IDENTITY NUC

Gapop 10.0, Gapept 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PC1\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PC1US\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.2	76.8	7120	9	US-09-380-773-2
2	18.6	74.4	200	9	US-09-969-373-877
3	18.6	74.4	221	9	US-09-754-853A-149
4	18.6	74.4	335	10	US-09-960-352-374
5	18.6	74.4	392	10	US-09-960-352-12239
6	18.6	74.4	469	9	US-09-918-995-9724
7	18.6	74.4	487	9	US-09-918-995-13475
8	18.6	74.4	1463	10	US-09-745-763-112
9	18.6	74.4	3003	12	US-10-044-205A-21
10	18.6	74.4	4020	9	US-10-198-846-15758
11	18.6	74.4	6071	9	US-10-239-676-78
12	18.6	74.4	335913	9	US-09-754-853A-2
13	18.6	74.4	335913	9	US-09-754-853A-3
14	18.2	72.8	132	10	US-09-969-373-868
15	18.2	72.8	234	9	US-10-066-543-2507
16	18.2	72.8	259	10	US-09-969-373-81
17	18.2	72.8	259	10	US-09-969-373-62
18	18.2	72.8	746	10	US-09-770-149-33
19	18.2	72.8	822	9	US-10-144-929-55

20	18.2	72.8	1410	9	US-09-925-299-179	Sequence 179, App
21	18.2	72.8	1410	9	US-10-106-698-1413	Sequence 1413, Ap
22	18.2	72.8	1410	10	US-09-925-299-179	Sequence 179, App
23	18.2	72.8	1755	9	US-10-106-698-363	Sequence 363, App
24	18.2	72.8	3015	9	US-10-128-714-147	Sequence 147, App
25	18.2	72.8	3015	9	US-10-128-714-5147	Sequence 5147, Ap
26	18.2	72.8	5775	9	US-09-866-557A-1	Sequence 1, Appl
27	18.2	72.8	5775	9	US-10-055-797-1	Sequence 1, Appl
28	18.2	72.8	11735	9	US-10-239-676-34	Sequence 34, Appl
29	18.2	72.8	335913	9	US-09-754-853A-2	Sequence 2, Appl
30	18.2	72.8	335913	9	US-09-754-853A-3	Sequence 3, Appl
31	18.2	72.8	337658	10	US-09-813-320-3	Sequence 3, Appl
32	18.2	71.2	369	10	US-09-864-761-13372	Sequence 13372, A
33	17.6	70.4	330	9	US-09-764-891-1640	Sequence 1640, Ap
34	17.6	70.4	339	9	US-09-764-891-7638	Sequence 7638, Ap
35	17.6	70.4	339	9	US-09-764-891-7639	Sequence 7639, Ap
36	17.6	70.4	339	9	US-09-764-891-7640	Sequence 7640, Ap
37	17.6	70.4	339	9	US-09-764-891-8289	Sequence 8289, Ap
38	17.6	70.4	339	9	US-09-764-891-8290	Sequence 8290, Ap
39	17.6	70.4	349	9	US-09-764-891-8291	Sequence 8291, Ap
40	17.6	70.4	349	9	US-09-813-358-180	Sequence 180, App
41	17.6	70.4	423	10	US-09-867-701-7081	Sequence 7081, Ap
42	17.6	70.4	442	9	US-09-796-692-663	Sequence 663, App
43	17.6	70.4	442	9	US-09-796-692-7547	Sequence 7547, Ap
44	17.6	70.4	442	9	US-09-796-692-7547	Sequence 7547, Ap
45	17.6	70.4	442	9	US-10-040-862-663	Sequence 663, App

#### ALIGNMENTS

```

RESULT 1
US-09-380-773-2/c
Sequence 2, Application US/09380773
Publication No. US20030113884A1
GENERAL INFORMATION:
APPLICANT: Hehn, Silke
APPLICANT: Schilling, Brigitte
APPLICANT: Gottschalk, Gerhard
TITLE OF INVENTION: Methods for the Biosynthesis of Polyesters
FILE REFERENCE: MOBT136--- 118899, 0136, NPUS00
CURRENT APPLICATION NUMBER: US/09/380,773
CURRENT FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: PCT/US97/03994
PRIOR FILING DATE: 1997-03-03
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 7120
TYPE: DNA
ORGANISM: Clostridium kluyveri
US-09-380-773-2

Query Match          76.8%; Score 19.2; DB 9; Length 7120;
Best Local Similarity 87.5%; Pred. No. 4.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 ccatatatactacattcaaaaca 24
        |||||
DB      3731 ccatatatactactccaaagca 3708

RESULT 2
US-09-969-373-877
Sequence 877, Application US/09969373
Patent No. US2002013852A1
GENERAL INFORMATION:
APPLICANT: Effertz, Roger J.
APPLICANT: Haugse, Brian M.
TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
FILE REFERENCE: 38-10(52679)A
CURRENT APPLICATION NUMBER: US/09/969,373

```



OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 251:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3652 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-251

Query Match 68.0%; Score 17; DB 4; Length 3652;  
Best Local Similarity 80.0%; Pred. No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 CCATATATCTCATTCATAACAA 25  
Db 2085 CCATAAATACCTGATTAAATCAA 2109

Search completed: July 8, 2003, 05:57:14  
Job time : 22.9965 secs



TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-09-054-298-17

Query Match 68.8%; Score 17.2; DB 3; Length 297;  
Best Local Similarity 86.4%; Pred. No. 1.4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ATATATATCTACATTCACAAACA 24  
|||||  
Db 267 ATATATATATATATACAAACA 288

RESULT 9  
US-08-818-655-17  
Sequence 17, Application US/08818655  
Patent No. 6258557

GENERAL INFORMATION:  
APPLICANT: Lee, Mu-En  
APPLICANT: Haber, Edgar  
APPLICANT: Jain, Mukesh  
APPLICANT: Yet, Shaw-Fang  
TITLE OF INVENTION: SMOOTH MUSCLE CELL LIM PROTEIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/818,655  
FILING DATE: 14-MAR-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/616,368  
FILING DATE: 15-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 05433/030001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 297 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-818-655-17

Query Match 68.8%; Score 17.2; DB 4; Length 297;  
Best Local Similarity 86.4%; Pred. No. 1.4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ATATATATCTACATTCACAAACA 24  
|||||  
Db 267 ATATATATATATATACAAACA 288

RESULT 10  
US-09-134-001C-2219/c  
Sequence 2219, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:

APPLICANT: Lynn Doucelte-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 2219  
LENGTH: 342  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2219

Query Match 68.8%; Score 17.2; DB 4; Length 342;  
Best Local Similarity 86.4%; Pred. No. 1.4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATTCACAA 22  
|||||  
Db 179 CCATCATATCTACATTCATTA 158

RESULT 11  
US-08-966-316-6/c  
Sequence 6, Application US/08966316  
Patent No. 5932445

GENERAL INFORMATION:  
APPLICANT: Lal, Preeti  
APPLICANT: Au-Young, Janice  
APPLICANT: Reddy, Roopa  
APPLICANT: Murthy, Lynn E.  
APPLICANT: Mathur, Preeti  
TITLE OF INVENTION: SIGNAL PEPTIDE - CONTAINING PROTEINS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/966,316  
FILING DATE: Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0424 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:

INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2111 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRSTNOT04

SEQ ID NO 1  
LENGTH: 11970  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-345-217-1

Query Match 70.4%; Score 17.6; DB 4; Length 11970;  
Best Local Similarity 83.3%; Pred. No. 93;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CATATATCTACATTCATAAACA 25  
DB 11895 CATACAGAGCTACATTAATAACA 11918

RESULT 6  
US-09-128-155-16  
Sequence 16, Application US/09128155  
Patent No. 6117654  
GENERAL INFORMATION:  
APPLICANT: Pan, Yang  
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY  
FILE REFERENCE: 09404/052001  
CURRENT APPLICATION NUMBER: US/09/128,155  
CURRENT FILING DATE: 1998-08-03  
EARLIER APPLICATION NUMBER: US 60/091,650  
EARLIER FILING DATE: 1998-07-02  
EARLIER APPLICATION NUMBER: US 60/054,646  
EARLIER FILING DATE: 1997-08-04  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 16  
LENGTH: 152331  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(152331)  
OTHER INFORMATION: n = A,T,C or G  
US-09-128-155-16

Query Match 70.4%; Score 17.6; DB 3; Length 152331;  
Best Local Similarity 83.3%; Pred. No. 88;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CATATATCTACATTCATAAACA 25  
DB 49288 CATACAGAGCTACATTAATAACA 49311

RESULT 7  
US-08-616-368A-17  
Sequence 17, Application US/08616368A  
Patent No. 5767262  
GENERAL INFORMATION:  
APPLICANT: Lee, Mu-En  
APPLICANT: Haber, Edgar  
APPLICANT: Jain, Mukesh  
APPLICANT: Yet, Shaw-Fang  
TITLE OF INVENTION: SMOOTH MUSCLE CELL LIM PROTEIN  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/616,368A  
FILING DATE: 15-MAR-1996  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 05433/022001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 297 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-616-368A-17

Query Match 68.8%; Score 17.2; DB 1; Length 297;  
Best Local Similarity 86.4%; Pred. No. 1,4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ATATATATCTACATTCATAACA 24  
DB 267 ATATATATATATATACATAACA 288

RESULT 8  
US-09-054-298-17  
Sequence 17, Application US/09054298  
Patent No. 6136953  
GENERAL INFORMATION:  
APPLICANT: Lee, Mu-En  
APPLICANT: Haber, Edgar  
APPLICANT: Jain, Mukesh  
APPLICANT: Yet, Shaw-Fang  
TITLE OF INVENTION: SMOOTH MUSCLE CELL LIM PROTEIN  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/054,298  
FILING DATE: 02-APR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/616,368  
FILING DATE: 15-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Beattie, Ph.D., Ingrid A.  
REGISTRATION NUMBER: P-42,306  
REFERENCE/DOCKET NUMBER: 05433/022002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 297 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,080
FILING DATE: 17-Jul-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,126
FILING DATE: 26-Jan-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0980
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 76:
SEQUENCE CHARACTERISTICS:
LENGTH: 1386 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: linear
TOPOLOGY: linear
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: INTRON 9 OF RAD50 GENOMIC SEQUENCE
US-08-687-080-76

Query Match          72.8%; Score 18.2; DB 2; Length 1386;
Best Local Similarity 87.0%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 CATATATCTACATCAAAACA 24
Db      278 CATAGAAATGTACATCAAAACA 300

RESULT 3
US-09-853-768-3
Sequence 3, Application US/09853768
Patent No. 6444466
GENERAL INFORMATION:
APPLICANT: Donna T. Ward
APPLICANT: Andrew T. Watt
TITLE OF INVENTION: ANTISENSE MODULATION OF HELICASE-NOI EXPRESSION
FILE REFERENCE: RUS-0217
CURRENT APPLICATION NUMBER: US/09/853,768
CURRENT FILING DATE: 2001-05-10
NUMBER OF SEQ ID NOS: 91
SEQ ID NO 3
LENGTH: 7037
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (183)...(5957)
US-09-853-768-3

Query Match          72.8%; Score 18.2; DB 4; Length 7037;
Best Local Similarity 87.0%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 ATATATCTACATCAAAACA 25
Db      1042 ATATATCTGTCATCAAAAGAA 1064

RESULT 4
US-08-487-826B-13
Sequence 13, Application US/08487826B
Patent No. 5993827
```

```
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Williams, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbé Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match          72.8%; Score 18.2; DB 2; Length 19124;
Best Local Similarity 87.0%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 CATATATCTACATCAAAACA 24
Db      7135 CATATATATATACATCAAAACA 7157

RESULT 5
US-09-345-217-1
Sequence 1, Application US/09345217
Patent No. 6268142
GENERAL INFORMATION:
APPLICANT: DUFF, GORDON W.
APPLICANT: COX, ANGELA
APPLICANT: CAMP, NICOLA J.
APPLICANT: DIGIOVINE, FRANCESCO S.
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR DISEASES ASSOCIATED
WITH AN IL-1 INFLAMMATORY HAPLOTYPE
FILE REFERENCE: MSA-010.02
CURRENT APPLICATION NUMBER: US/09/345,217
CURRENT FILING DATE: 1999-06-30
EARLIER APPLICATION NUMBER: PCT/GB98/01481
EARLIER FILING DATE: 1998-05-21
EARLIER APPLICATION NUMBER: 9711040.7
EARLIER FILING DATE: 1997-05-29
NUMBER OF SEQ ID NOS: 32
SOFTWARE: Patentin Ver. 2.0
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:45:20 / Search time 5.99652 Seconds  
(without alignments)  
1278.561 Million cell updates/sec

Title: US-09-477-082-32

Sequence: 1 ccatatatactacattcaaaaacaa 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents, NA: \*  
1: /cgn2\_6/prodata/1/lna/5A\_COMB.seq: \*  
2: /cgn2\_6/prodata/1/lna/5B\_COMB.seq: \*  
3: /cgn2\_6/prodata/1/lna/6A\_COMB.seq: \*  
4: /cgn2\_6/prodata/1/lna/6B\_COMB.seq: \*  
5: /cgn2\_6/prodata/1/lna/PCUTS\_COMB.seq: \*  
6: /cgn2\_6/prodata/1/lna/Backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	18.6	74.4	798	4	US-09-280-116-82
2	18.2	72.8	1386	2	US-08-687-080-76
3	18.2	72.8	7037	4	US-09-853-768-3
4	18.2	72.8	19124	2	US-08-487-826B-13
5	17.6	70.4	11970	4	US-09-345-217-1
6	17.6	70.4	152331	3	US-09-128-155-16
7	17.2	68.8	297	3	US-08-616-368A-17
8	17.2	68.8	297	3	US-09-054-298-17
9	17.2	68.8	297	3	US-08-818-655-17
10	17.2	68.8	342	4	US-09-134-001C-2219
11	17.2	68.8	2111	2	US-08-966-316-6
12	17.2	68.8	945	4	US-08-858-207A-111
13	17.2	68.8	1952	4	US-09-315-444-115
14	17.2	68.8	1952	4	US-09-721-362-115
15	17.2	68.8	3652	4	US-08-961-527-251
16	17.2	68.8	10754	2	US-08-966-958-1
17	17.2	68.8	10754	2	US-09-315-817-1
18	17.2	68.8	10754	2	US-09-342-353-1
19	17.2	68.8	57	4	US-08-273-594-33
20	16.6	66.4	57	3	US-09-094-919-33
21	16.6	66.4	1260	3	US-08-793-331-1
22	16.6	66.4	2763	4	US-09-489-868A-3
23	16.6	66.4	3023	4	US-09-203-453-4
24	16.6	66.4	3959	1	US-08-474-067-1
25	16.6	66.4	3959	2	US-08-474-068A-1
26	16.6	66.4	3959	2	US-08-472-481-1
27	16.6	66.4	5622	4	US-09-067-800-3

C 28	16.6	66.4	5622	4	US-09-349-677-3	Sequence 3, Appl
C 29	16.6	66.4	10007	4	US-09-410-464-13	Sequence 13, Appl
C 30	16.6	66.4	70000	4	US-09-851-896-3	Sequence 3, Appl
C 31	16.6	66.4	72604	4	US-09-268-992-7	Sequence 7, Appl
C 32	16.6	66.4	72604	4	US-09-657-474-7	Sequence 7, Appl
C 33	16.6	66.4	246240	2	US-08-724-394A-20	Sequence 20, Appl
C 34	16.6	66.4	246240	2	US-08-724-394A-21	Sequence 21, Appl
C 35	16.6	66.4	246240	2	US-08-724-394A-22	Sequence 22, Appl
C 36	16.4	65.6	435	4	US-09-134-001C-1206	Sequence 1206, Ap
C 37	16.4	65.6	10614	1	US-08-135-511-35	Sequence 35, Appl
C 38	16.4	65.6	10614	1	US-08-187-453-35	Sequence 35, Appl
C 39	16.2	64.8	470	4	US-09-221-017B-324	Sequence 324, App
C 40	16.2	64.8	685	4	US-08-822-999-10	Sequence 10, Appl
C 41	16.2	64.8	693	2	US-08-791-347-16	Sequence 16, Appl
C 42	16.2	64.8	2000	4	US-09-056-285A-3	Sequence 3, Appl
C 43	16.2	64.8	2004	4	US-08-961-527-95	Sequence 95, Appl
C 44	16.2	64.8	2606	4	US-09-252-816A-7	Sequence 7, Appl
C 45	16.2	64.8	2981	1	US-08-257-073-2	Sequence 2, Appl

#### ALIGNMENTS

RESULT 1  
US-09-280-116-82  
Sequence 82, Application US/09280116A  
Patent No. 6331427  
GENERAL INFORMATION:  
APPLICANT: Robison, Keith E.  
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs  
FILE REFERENCE: 5800-24, 035800/176965  
CURRENT APPLICATION NUMBER: US/09/280,116A  
CURRENT FILING DATE: 1999-03-26  
NUMBER OF SEQ ID NOS: 268  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 82  
LENGTH: 798  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: matrix metalloproteases  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1)-(798)  
OTHER INFORMATION: n = a, t, c or g  
US-09-280-116-82

Query Match 74.4%; Score 18.6; DB 4; Length 798;  
Best Local Similarity 84.0%; Pred. No. 40;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCAATATATCTCATTCACAAACAA 25  
||||||| || ||| |||  
DB 422 CCAATATATATATCAACACAA 446

RESULT 2  
US-08-687-080-76  
Sequence 76, Application US/08687080  
Patent No. 5965427  
GENERAL INFORMATION:  
APPLICANT: Gregory Dolganov  
TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof  
NUMBER OF SEQUENCES: 175  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

REFERENCE 3 (bases 1 to 923)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequençage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvareum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES  
Source  
Location/Qualifiers  
1..923  
/organism="Pichia farinosa"  
/strain="CBS 7064"  
/db\_xref="taxon:4920"  
/clone="AX0AA030H08"  
/clone\_1bp="AX0AA"  
/note="end : T3"  
misc\_feature  
10..>150  
/note="similar to *Saccharomyces cerevisiae* ORF YBR217w [APC42 : component of the autophagic system ]"  
misc\_feature  
complement(652..>918)  
/note="similar to *Saccharomyces cerevisiae* ORF YIL123w [SIM1 : involved in cell cycle regulation and aging ] similar to *Saccharomyces cerevisiae* ORF YKR42w [ UTH1 : involved in the aging process ] similar to *Saccharomyces cerevisiae* ORF YNL066w [ SUM4 : cell wall protein with homology to beta-glucosidase ] evidence=not\_experimental  
misc\_feature  
complement(676..>915)  
/note="similar to *Saccharomyces cerevisiae* ORF YUL16c [ NCAs : involved in regulation of synthesis of Atppp and Atppp ]"  
BASE COUNT 326 a 203 c 156 g 236 t 2 others  
ORIGIN  
Query Match 77.6%; Score 19.4; DB 17; Length 923;  
Best Local Similarity 95.2%; Fred. No. 1.4e-03;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

cy 3 ATATATATCTACATCAAAAC 23  
||||| |||||||||  
Db 28 ATATATGCTCTACATTCAAAAC 48

RESULT 15  
CNS06MCW 965 bp DNA linear GSS 06-JUL-2001  
LOCUS T3 end of clone AX0AA030D10 of library AX0AA from strain CBS 7064  
DEFINITION of *Pichia farinosa*, genomic survey sequence.  
ACCESSION AL418278  
VERSION AL418278.1 GI:12200835  
KEYWORDS GSS.  
SOURCE *Pichia farinosa*.  
ORGANISM *Pichia farinosa*.  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; *Pichia*.  
REFERENCE 1 (bases 1 to 965)  
Soudier,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
Boilotin-Pikuhara,M., Bon,E., Brothier,P., Casaregola,S.,  
deMontigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,  
Malpertuy,A., Neuvéglise,C., Olier-Kalogeropoulos,O., Potier,S.,  
Saurin,W., Tekaia,F., Toffano-Niccho,C., Wesolowski-Louvel,M.,  
Winkler,P. and Weissenbach J.  
Genomic exploration of the hemiascomycetous yeasts: 1. A set of

```

JOURNAL MEDLINE 20584711 Yeast species for molecular evolution studies
JOURNAL MEDLINE 11152876 FERS Lett. 487 (1), 3-12 (2000)
REFERENCE 2 (bases 1 to 965)
AUTHORS de Montigny,J., Spehner,C., Souciet,J., Tekala,F., Dujon,B.,
Winkler,P., Artiguenave,F. and Poulter,S.
TITLE Genomic exploration of the hemiascomycetous yeasts: 15. Pichia
sorbitophila
JOURNAL MEDLINE 20584725 FERS Lett. 487 (1), 87-90 (2000)
JOURNAL MEDLINE 11152890
REFERENCE 3 (bases 1 to 965)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segre@genoscope.cns.fr Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a Random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
Source Location/Qualifiers
misc_feature
1..965
/organism="Pichia farinosa"
/strain="CBS 7064"
/db_xref="taxon:4920"
/clone="AX0A030D10"
/clone_1lp="AX0A"
/note="end : T3"
<10..>150
/note="similar to Saccharomyces cerevisiae ORF YBR217w [
APG32 ; component of the autophagic system ]"
evidence=not_experimental
misc_feature
1..962
/note="similar to Saccharomyces cerevisiae ORF YJL123w [
SIM1 ; involved in cell cycle regulation and aging ]
similar to Saccharomyces cerevisiae ORF YKR042w [ UTH1 ;
involved in the aging process ]
similar to Saccharomyces cerevisiae ORF YNL066w [ SUN4 ;
cell wall protein with homology to beta-glucosidase ]"
evidence=not_experimental
misc_feature
1..942
/note="similar to Saccharomyces cerevisiae ORF YJL16c [
NCA3 ; involved in regulation of synthesis of Atp6p and
Atbp ]"
evidence=not_experimental
BASE COUNT 339 a 214 c 164 g 247 t 1 others
ORIGIN
Query Match 77.6% Score 19.4; DB 17; Length 965;
Best Local Similarity 95.2% ; Pred. No. 1.4e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0.
3 ATATATATCATCATCAAAAC 23
|||||
DB 28 ATATATGTCATCATCAAAAC 48

```

REFERENCE 1 (bases 1 to 651)  
AUTHORS Zhao, S., Adams, M.D., Niernan, W., Malek, J., de Jong, P., and Venter, J.C.  
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building  
JOURNAL Unpublished (1997)  
COMMENT Other\_GSSs: RPCI11-12117.TV  
Contact: Shaying Zhao, William Niernan, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hdebtigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACpac Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (<http://info@resgen.com>). BAC end search page: [http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: SP6  
Class: BAC ends.  
FEATURES  
Source Location/Qualifiers  
1..651  
/organism="Homo sapiens"  
/db\_xref="GDB:7546278"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-12117"  
/clone\_11b="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBAC3.6, Site\_1: EcoRI; Site\_2: EcoRI; RPCI11 Human Male BAC Library"  
BASE COUNT 213 a 81 c 96 g 261 t  
ORIGIN  
Query Match 77.6%; Score 19.4; DB 17; Length 651;  
Best Local Similarity 95.2%; Pred. No. 1.4e+03;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3 ATATATCTACATTCATAAC 23  
|||||  
Db 296 ATATATGTACATTCATAAC 276  
|||||  
RESULT 13  
CNS060RT 849 bp DNA linear GSS 06-JUL-2001  
LOCUS T3 end of clone AX0AA012F02 of library AX0AA from strain CBS 7064  
DEFINITION of Pichia farinosa, genomic survey sequence.  
ACCESSION AL416212  
VERSION AL416212.1 GI:12196387  
KEYWORDS GSS.  
SOURCE Pichia farinosa.  
ORGANISM Pichia farinosa.  
REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.  
1 (bases 1 to 849)  
AUTHORS Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Ben, E., Brothier, P., Casaregola, S., de Montigny, J., Dujon, B., Durans, P., Lepingle, A., Llorente, B., Malpertuy, A., Nevegilise, C., Ozler-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P., and Weissenbach, J.  
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies  
JOURNAL FEMS Lett. 487 (1), 3-12 (2000)  
MEDLINE 20584711  
PUBMED 11152876  
REFERENCE 2 (bases 1 to 849)  
AUTHORS de Montigny, J., Spehner, C., Souciet, J., Tekala, F., Dujon, B., Wincker, P., Artiguenave, F., and Potier, S.

TITLE Genomic exploration of the hemiascomycetous yeasts: 15. Pichia sorbitophila  
JOURNAL FEMS Lett. 487 (1), 87-90 (2000)  
MEDLINE 20584725  
PUBMED 11152890  
REFERENCE 3 (bases 1 to 849)  
AUTHORS Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Ben, E., Brothier, P., Casaregola, S., de Montigny, J., Dujon, B., Durans, P., Lepingle, A., Llorente, B., Malpertuy, A., Nevegilise, C., Ozler-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P., and Weissenbach, J.  
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies  
JOURNAL FEMS Lett. 487 (1), 3-12 (2000)  
MEDLINE 20584711  
PUBMED 11152876  
REFERENCE 2 (bases 1 to 923)  
AUTHORS de Montigny, J., Spehner, C., Souciet, J., Tekala, F., Dujon, B., Wincker, P., Artiguenave, F., and Potier, S.  
TITLE Genomic exploration of the hemiascomycetous yeasts: 15. Pichia sorbitophila  
JOURNAL FEMS Lett. 487 (1), 87-90 (2000)  
MEDLINE 20584725  
PUBMED 11152890  
FEATURES  
Source Location/Qualifiers  
1..849  
/organism="Pichia farinosa"  
/strain="CBS 7064"  
/db\_xref="taxon:4920"  
/clone="AX0AA012F02"  
/clone\_11b="AX0AA"  
/note="end : T3"  
BASE COUNT 280 a 167 c 150 g 242 t 10 others  
ORIGIN  
Query Match 77.6%; Score 19.4; DB 17; Length 849;  
Best Local Similarity 95.2%; Pred. No. 1.4e+03;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3 ATATATCTACATTCATAAC 23  
|||||  
Db 553 ATATATGTACATTCATAAC 573  
|||||  
RESULT 14  
CNS060MET 923 bp DNA linear GSS 06-JUL-2001  
LOCUS T3 end of clone AX0AA030H08 of library AX0AA from strain CBS 7064  
DEFINITION of Pichia farinosa, genomic survey sequence.  
ACCESSION AL418347  
VERSION AL418347.1 GI:12200957  
KEYWORDS GSS.  
SOURCE Pichia farinosa.  
ORGANISM Pichia farinosa.  
REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.  
1 (bases 1 to 923)  
AUTHORS Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Ben, E., Brothier, P., Casaregola, S., de Montigny, J., Dujon, B., Durans, P., Lepingle, A., Llorente, B., Malpertuy, A., Nevegilise, C., Ozler-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P., and Weissenbach, J.  
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies  
JOURNAL FEMS Lett. 487 (1), 3-12 (2000)  
MEDLINE 20584711  
PUBMED 11152876  
REFERENCE 2 (bases 1 to 923)  
AUTHORS de Montigny, J., Spehner, C., Souciet, J., Tekala, F., Dujon, B., Wincker, P., Artiguenave, F., and Potier, S.  
TITLE Genomic exploration of the hemiascomycetous yeasts: 15. Pichia sorbitophila  
JOURNAL FEMS Lett. 487 (1), 87-90 (2000)  
MEDLINE 20584725  
PUBMED 11152890



FEATURES  
source  
Location/Qualifiers  
1. 764  
/organism="Brassica oleracea"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"  
/clone\_lib="BO\_2\_3\_KB"  
/note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT  
223 a 140 c 171 g 230 t

ORIGIN

Query Match  
Best Local Similarity 91.3%; Pred. No. 1e+03;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 ATATATCTACATTCAAAACA 25  
|||||  
Db 723 ATATATCTACATTCAAAATA 745

RESULT 10  
CNS04K6T/c  
LOCUS  
DEFINITION  
Tetradodon nigroviridis genome survey sequence T7 end of clone 115P02 of library G from Tetradodon nigroviridis, genomic survey sequence.  
ACCESSION  
AL294446.1 GI:8033026  
KEYWORDS  
GSS; genome survey sequence.  
SOURCE  
Tetradodon nigroviridis.  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetradodon.  
1 (bases 1 to 793)  
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissenbach,J.  
Human gene number estimate provided by genome wide analysis using Tetradodon nigroviridis DNA sequence

TITLE  
Unpublished  
2 (bases 1 to 793)  
Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.  
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradodon nigroviridis

JOURNAL  
Unpublished  
3 (bases 1 to 793)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000)  
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetradodon>.

FEATURES  
source  
Location/Qualifiers  
1. 793  
/organism="Tetradodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone\_lib="115P02"  
/clone\_lib="G"  
/note="Genoscope sequence ID: COBG115DH01.P1-end : T7"

BASE COUNT  
220 a 94 c 208 g 240 t 31 others

ORIGIN

Query Match  
Best Local Similarity 84.0%; Pred. No. 1e+03;  
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 CCATATATCTACATTCAAAACA 25

Db 198 CCATATATCTCTCATTMAAAACA 174  
|||||  
RESULT 11  
A2519839/c  
LOCUS  
DEFINITION  
RPCI-11-79F2.TVB RPCI-11 Homo sapiens genomic clone RPCI-11-79F2, DNA sequence.  
ACCESSION  
A2519839.1 GI:10831743  
KEYWORDS  
GSS.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 513)  
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.  
BAC end sequences of library RPCI-11  
Unpublished (1997)  
Other GSSs: RPCI11-79F2.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.bufileo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.bufileo.edu/ordering>) or from Research Genet cs (<http://infogenet.com>). BAC end search page: [http://www.tigr.org/cdb/human/bac/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/cdb/human/bac/bac_end_search/bac_end_search.html). This BAC end was generated during the Rdb process and may have higher chance of clone tracking errors.  
Seq primer: T7  
Class: BAC ends.

FEATURES  
source  
Location/Qualifiers  
1. 513  
/organism="Homo sapiens"  
/db\_xref="GDB:7530873"  
/db\_xref="taxon:9606"  
/clone\_lib="RPCI-11-79F2"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPCI11 Human Male BAC Library"

BASE COUNT  
168 a 65 c 76 g 203 t 1 others

ORIGIN

Query Match  
Best Local Similarity 77.6%; Score 19.4; DB 17; Length 513;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 ATATATCTACATTCAAAAC 23  
|||||  
Db 269 ATATATCTACATTCAAAAC 249

RESULT 12  
AQ342956/c  
LOCUS  
DEFINITION  
RPCI11-12117.TU RPCI-11 Homo sapiens genomic clone RPCI-11-12117, DNA sequence.  
ACCESSION  
AQ342956.1 GI:4167852  
KEYWORDS  
GSS.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Otsuka Pharmaceutical Co., Ltd  
463-10 Kagasuno Kawasuchi-cho, Tokushima, Tokushima, 771-01 Japan  
Tel: 0886-65-2888  
Fax: 0886-37-1035.

FEATURES  
source  
Location/Qualifiers  
1..369  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Clontech human aorta polyA+ mRNA (#6572)"  
/note="Male adult, hematopoietic tissue, stem cell"

BASE COUNT  
ORIGIN  
108 a 54 c 57 g 140 t 10 others

Query Match 79.2%; Score 19.8; DB 14; Length 369;  
Best Local Similarity 84.0%; Pred. No. 1e+03;  
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCATATATCTACATCAACA 25  
Db 79 CCATATATATATACAAACA 103

RESULT 7  
AA966215/c  
LOCUS  
DEFINITION  
v4e09a1.f1 Aspergillus nidulans 24hr asexual developmental and  
vegetative cDNA lambda zap library Emericella nidulans cDNA clone  
v4e09a1 3', mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AA966215.1 GI:3140297  
EST.  
Emericella nidulans.  
Emericella nidulans.  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiiales; Trichocomaceae; Emericella.  
1 (bases 1 to 371)  
Kupfer, D., Gray, J., Hausner, J., Lai, H., Martin, W., Aramayo, R.,  
Prade, R. and Roe, B.  
An Aspergillus nidulans EST Database  
Unpublished (1998)  
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu  
Department of Chemistry and Biochemistry  
Advanced Center for Genome Technology, University of Oklahoma  
620 Partington Oval, Norman, OK 73019, USA  
Tel: 405 325 4912  
Fax: 405 325 7762  
Email: broe@ou.edu

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
We anticipate the future release of the cDNA clones to the Fungal  
Genetics Stock Center  
Seq primer: M13-20  
High quality sequence stop: 208.

FEATURES  
source  
Location/Qualifiers  
1..371  
/organism="Emericella nidulans"  
/strain="FGSC A26"  
/db\_xref="taxon:162425"  
/clone\_lib="v4e09a1"  
/clone\_lib="Aspergillus nidulans 24hr asexual  
developmental and vegetative cDNA lambda zap library"  
/tissue\_type="vegetative mycelia, asexual structures"  
/note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; 5' end of cDNA cloned into EcoRI site of Bluescript  
3' end of cDNA cloned into XhoI site of Bluescript"

BASE COUNT  
ORIGIN  
129 a 60 c 106 g 76 t

Query Match 79.2%; Score 19.8; DB 9; Length 371;  
Best Local Similarity 91.3%; Pred. No. 1e+03;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CATATATCTACATCAACA 24  
||||||| ||||| |||||

Db 47 CATATATCTACATCAACA 25

RESULT 8  
Bt401889/c

LOCUS  
DEFINITION  
Bt401889 Dictyostelium discoideum cDNA library, SF Dictyostelium  
discoideum cDNA clone dds19a20 3', mRNA sequence.  
Bt401889  
Bt401889.1 GI:19314806

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
EST.  
Dictyostelium discoideum.  
Dictyostelium discoideum.  
Eukaryota; Mycetozoa: Dictyostellida; Dictyostelium.  
1 (bases 1 to 614)  
Urushihara, H., Tanaka, Y., Kohara, Y. and Shin, I. T.  
Full length cDNA of Dictyostelium discoideum at the slug stage  
Unpublished (2002)  
Contact: Tadasi Shin-I  
Center for Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tsuhit@genes.nig.ac.jp.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Location/Qualifiers  
1..614  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="dds19a20"  
/clone\_lib="Dictyostelium discoideum cDNA library, SF"  
/sex="mat A"  
/dev\_stage="Slug stage"

BASE COUNT  
ORIGIN  
244 a 65 c 70 g 234 t 1 others

Query Match 79.2%; Score 19.8; DB 13; Length 614;  
Best Local Similarity 91.3%; Pred. No. 1e+03;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATATATCTACATCAACA 25  
Db 503 ATATATCTACATCAACA 481

RESULT 9  
BH701475  
LOCUS  
DEFINITION  
BH701475 764 bp DNA linear GSS 20-FEB-2002  
BOMH023TF BO\_2\_3\_KB Brassica oleracea genomic clone BOMH023, DNA  
sequence.  
BH701475  
BH701475.1 GI:18779250  
GSS.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
BH701475  
GSS.  
Brassica oleracea.  
Brassicaceae.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
1 (bases 1 to 764)  
Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other-GSSs: BOMH023TR  
Contact: Chris Town  
TIGR

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
7912 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TF  
Class: sheared ends.

LOCUS	DRLP18T	582 bp	DNA	linear	GSS 06-JUN-2002
DEFINITION	Danio rerio genomic clone Dkey-1P18, genomic survey sequence.				
ACCESSION	AL737904				
VERSION	AL737904.1				
KEYWORDS	GSS.				
SOURCE	zebrafish.				
ORGANISM	Danio rerio				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Actinopterygii; Neopterygii; Teleostei; Ostariophysi;				
TITLE	Cypriniformes; Cyprinidae; Danio.				
JOURNAL	1 (bases 1 to 582)				
	Humphray,S.J., Huckle,E. and Hunt,S.E.				
	Direct Submission				
	Submitted (06-JUN-2002) The Sanger Institute, Wellcome Trust Genome				
	Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:				
	humphrays@sanger.ac.uk Unpublished				
	This sequence was generated from the T7 end of BAC 1P18, 1P18 is				
	part of the Daniokey Pilot BAC Library created by R. Plasterk and				
	N.V. Keygene.				
COMMENT	Further details: http://www.sanger.ac.uk/Projects/D_rerio/.				
FEATURES	Location/Qualifiers				
SOURCE	1..582				
	/organism="Danio rerio"				
	/db_xref="taxon:7955"				
	/clone="Dkey-1P18"				
	/rname_type="Testis"				
	/note="vector pindigobAC-536"				
BASE COUNT	191 a 86 c 101 g 204 t				
ORIGIN					
Query Match	80.8%; Score 20.2; DB 17; Length 582;				
Best Local Similarity	88.0%; Pred. No. 7.1e+02;				
Matches	22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;				
OY	1 CCATATATCTCATCTCAACAACAA 25				
Db					
	37 CCACATATATTGCATTCACAACAA 13				
RESULT 5					
BB404480					
DEFINITION	BB404480 RIKEN full-length enriched, ES cells Mus musculus cDNA				
ACCESION	BBA04480				
VERSION	BBA04480.1 GI:9223876				
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muride; Murinae; Mus.				
REFERENCE	1 (bases 1 to 318)				
AUTHORS	Kono,H., Aitawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci				
	, P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,				
	Hirozane,T., Horii,F., Ishii,Y., Ishikawa,T., Ishikawa,T., Itoh,M.,				
	Iizawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,				
	Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,				
	Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.				
	, Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shiba				
	, Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,T., Suganaka,T.				
	, Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya				
	, T., Tsunoda,Y., Watabiki,A., Watanabe,S., Yamamura,T., Yamataka,I.				
	, Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino				
	, M., Muramatsu,M. and Hayashizaki,Y.				
	RIKEN Mouse ESTs (Kono,H., et al.)				
TITLE	Unpublished (2000)				
JOURNAL	Contact: Yoshihide Hayashizaki				
COMMENT	Laboratory for Genome Exploration Research Group, RIKEN Genomic				
	Sciences Center(GSC), Yokohama Institute				
	The Institute of Physical and Chemical Research (RIKEN)				
	1-7-22 Shibariko-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan				
	Tel.: 81-45-503-9222				
	Fax: 81-45-503-9216				

Email: genome-res@cc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P. Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasakawa, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Thermosensitization and thermoinactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Iwama, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

Location/Qualifiers

1. 318  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="C330036108"  
/clone\_lib="RIKEN full-length enriched, ES cells"  
/cell\_type="ES cells"  
/lab\_host="SODR"  
/note="Site\_1: XhoI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'  
GAGGAGAGAGAGATCCACAGACGCTTTTCTTTTCTTTTCTTTVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5'  
GAGGAGAGAGATCTCGAGATTAAATTAATATCCCCCCCCCC 3']."

BASE COUNT 96 a 55 c 39 g 128 t

ORIGIN

Query Match 79.2%; Score 19.8; DB 10; Length 318;  
Best Local Similarity 91.3%; Pred. No. 1e+03; 2; Indels 0; Gaps 0;  
Matches 21; Conservative 0; Mismatches 2;

QY 3 ATATATATCTACATTCAAACAA 25  
|||||  
DB 220 ATATTATCTACATTCATTAACAA 242

RESULT 6  
D62159 369 bp mRNA linear EST 29-AUG-1999  
DEFINITION HW2424G12B Clontech human aorta poly(A<sup>+</sup>) mRNA (#6572) Homo sapiens  
ACCESSION D62159  
VERSION D62159.1 GI:965935  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 369)  
Fujiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M., Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H., Takaichi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y., Maekawa, H., Shin, S. and Nakamura, Y. Fujiwara et al. (1995)  
Unpublished (1995)  
Contact: Tsutomu Fujiwara  
Otsuka Gen Research Institute





PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259676.

## (HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI: 2001-483426/52.

P-PSDB; AAM88893.

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -

PS Claim 1; SEQ ID NO 6734; 3071pp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAM62170 to AAM91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
 CC represent sequences used in the exemplification of the present invention.

XX Sequence 379 BP; 108 A; 62 C; 104 G; 101 T; 4 other;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 CCATATATACCTACATTCAAACAA 25  
 ||| ||||| ||||| |||||  
 Db 209 CCACATATATCTTCATTAAAAACCA 185  
 Search completed: July 8, 2003, 03:11:19  
 Job time : 27.9421 secs

Query Match 74.4%; Score 18.6; DB 22; Length 379;  
 Best Local Similarity 84.0%; Pred. No. 3.3e+02;

CC The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA03910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapies. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoietic regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 361 BP; 100 A; 52 C; 73 G; 136 T; 0 other;

Query Match 74.4%; Score 18.6; DB 22; Length 361;  
Best Local Similarity 84.0%; Pred. No. 3.3e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 CCATATATCTACATCAACAA 25  
175 CTATATATATTTCATTCAAAAA 151

## RESULT 15

AAK61674/c  
ID AAK61674 standard; cDNA; 379 BP.

XX AAK61674;

DT 06-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:6734.

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytosolic; gene therapy; vaccine; metastasis; ss.

OS Homo sapiens.

PN WO200157182-A2.

XX 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US01354.

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214866.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 11-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227209.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234597.

PR 25-SEP-2000; 2000US-0234598.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235634.

PR 27-SEP-2000; 2000US-0235635.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239335.

PR 13-OCT-2000; 2000US-0239337.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.

PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.

KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIC-) EPIDENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI: 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful  
 for diagnosis and treatment of diseases associated with abnormal

PT cytosine methylation

PS Claim 1; SEQ ID NO 35; 32bp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/neurodegenerative bowel  
 CC diseases. The present sequence is a gene of the invention.

CC Sequence 9905 BP; 3133 A; 75 C; 1920 G; 4777 T; 0 other;

Query Match 75.2%; Score 18.8; DB 24; Length 9905;

Best Local Similarity 90.9%; Pred. No. 3e+02; Mismatches 2; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 ATATATATCTACTTCAAAACA 24

DB 7931 ATATATATATACATCAAAACA 7910

RESULT 13

AA161518 ID AA161518 standard; DNA; 221 BP.

AC AA161518;

DT 16-OCT-2001 (first entry)

DE Soybean 240017 region G3 DNA, SEQ ID NO: 149.

KW Soybean; antihelminthic; gene therapy; soybean cyst nematode; SCN;

KW SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;

KW 240017 region G3; 318013 region A3; 515002 region G2; ds.

OS Glycine max.

PN WO200151627-A2.

PD 19-JUL-2001.

PF 05-JAN-2001; 2001WO-US00552.

PR 07-JAN-2000; 2000US-0174880.

PA (MONS) MONSANTO CO.  
 PI Haughe BM, Wang ML, Parsons JD, Farnell LD;  
 DR WPI: 2001-425872/45.

PT New purified nucleic acid for producing a soybean plant having soybean  
 PT cyst nematode resistance and for use in plant breeding programs -  
 PS Claim 25; Page 1062; 1353pp; English.

CC The invention relates to nucleic acid molecules from regions of the  
 CC soybean genome which are associated with soybean cyst nematode (SCN)  
 CC resistance. The nucleic acids are used to transform plants, and can  
 CC produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.  
 CC The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes  
 CC of soybean plants and for introgressing SCN resistance or partial SCN  
 CC resistance into soybean plants. They can also be used in plant breeding  
 CC programmes. The invention also relates to proteins encoded by such  
 CC nucleic acid molecules, as well as antibodies capable of recognising  
 CC these proteins. The present sequence is a nucleic acid molecule  
 CC provided in the specification.

CC Sequence 221 BP; 72 A; 28 C; 28 G; 93 T; 0 other;

Query Match 74.4%; Score 18.6; DB 22; Length 221;

Best Local Similarity 84.0%; Pred. No. 5.3e+02; Mismatches 4; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 CCATATATATCTACTTCAAAACA 25

DB 120 CGATATATATATACATAAAACA 144

RESULT 14

AA191681/C ID AA191681 standard; CDNA; 361 BP.

AC AA191681;

DT 06-NOV-2001 (first entry)

DE Human polynucleotide SEQ ID NO 11741.

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation; ss.

OS Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US04927.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI: 2001-514838/56.

DR P-PSDB; AAC11750.

PT Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 PS Claim 1; SEQ ID NO 11741; 1399pp + Sequence Listing; English.



Query Match 76.8%; Score 19.2; DB 24; Length 15373;  
 Best Local Similarity 87.5%; Pred. No. 2.1e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATATATCTCATCTCAAAACAA 25  
 DB 12222 CTTAAATATCTCATCTCAAAATAA 12199

RESULT 10  
 ABL33519/C  
 ID ABL33519 standard; DNA: 5520 BP.

AC ABL33519;  
 XX  
 DT 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 1492.

XX Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianaemic; cytosatic; noctropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antinflamatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.

XX Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPig-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful  
 for diagnosis and treatment of diseases associated with abnormal  
 cytosine methylation

PS Claim 1; SEQ ID NO 1492; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.

SQ Sequence 5520 BP; 1725 A; 36 C; 1004 G; 2753 T; 2 other;

Query Match 75.2%; Score 18.8; DB 24; Length 5520;  
 Best Local Similarity 90.9%; Pred. No. 2.9e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 TATATATCTCATCTCAAAACAA 25  
 DB 2604 TAAATATCTCATCTCAAAACAA 2583

RESULT 11  
 ABL32705/C

ID ABL32705 standard; DNA: 6297 BP.

XX ABL32705;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 678.

XX Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianaemic; cytosatic; noctropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antinflamatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.

XX Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPig-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful  
 for diagnosis and treatment of diseases associated with abnormal  
 cytosine methylation

PS Claim 1; SEQ ID NO 678; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.

SQ Sequence 6297 BP; 1577 A; 217 C; 1471 G; 3032 T; 0 other;

Query Match 75.2%; Score 18.8; DB 24; Length 6297;  
 Best Local Similarity 90.9%; Pred. No. 3e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 TATATATCTCATCTCAAAACAA 25  
 DB 480 TAAATATCTCATCTCAAAACAA 459

RESULT 12

ABL32062/C

ID ABL32062 standard; DNA: 9905 BP.

AC ABL32062;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 35.

XX Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianaemic; cytosatic; noctropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 XX  
 SQ Sequence 4654 BP; 1313 A; 50 C; 955 G; 2336 T; 0 other;  
 Query Match 76.8%; Score 19.2; DB 24; Length 4654;  
 Best Local Similarity 87.5%; Pred. No. 2.1e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 CATATATATCTACATCAAAACAA 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 874 CATATTATCTACATCAAAACAA 851

RESULT 8  
 AAS46544/c  
 ID AAS46544 standard; DNA; 8845 BP.  
 AC AAS46544;  
 XX  
 AC AAS46544;  
 XX  
 DT 18-DEC-2001 (first entry)  
 DE Tumour suppressor gene derived chemically modified sequence #266.  
 XX  
 KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;  
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
 KW cytosine methylation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200168912-A2.  
 XX  
 PD 20-SEP-2001.  
 XX  
 PE 15-MAR-2001; 2001WO-EP02955.  
 XX  
 PE 15-MAR-2000; 2000DE-1013847.  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 DR WPI; 2001-602752/68.  
 XX  
 PT Fragments of chemically modified genes associated with tumour suppressor  
 PT genes and oncogenes, useful in designing primers and probes for  
 PT analysing diseases associated with cytosine methylation state e.g.  
 PT cancer -  
 XX  
 PS Claim 1; SEQ ID No 266; 27pp; English.  
 XX  
 CC The invention relates to a nucleic acid comprising a sequence of 18  
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
 CC bisulphite, of genes associated with tumour suppression and  
 CC oncogenes having a sequence taken from 536 (actually 533 since  
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a  
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
 CC form part of a set of probes for detecting the cytosine methylation state  
 CC and/or single nucleotide polymorphisms and also to be used in an  
 CC array for analysing diseases associated with CpG dinucleotides e.g.  
 CC cancers and tumours. The probes can also be used in a method for  
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
 CC and/or therapy of existing diseases or the predisposition to specific  
 CC diseases, by analysing cytosine methylations. The parameters may be  
 CC compared to another set of genetic and/or epigenetic parameters, the  
 CC differences serving as basis for diagnosis and/or prognosis events which  
 CC are disadvantageous to patients. The present sequence is one of the  
 CC 533 genomic sequences derived from tumour suppressor genes and

CC oncogenes. Sequences with even numbered Seq ID numbers are the  
 CC complementary sequence of the corresponding odd numbered sequence (e.g.  
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence  
 CC is missing).  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 8845 BP; 2565 A; 135 C; 1694 G; 4451 T; 0 other;  
 Query Match 76.8%; Score 19.2; DB 22; Length 8845;  
 Best Local Similarity 87.5%; Pred. No. 2.1e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 CATATATATCTACATCAAAACAA 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 8144 CAATATATCAACATCAAAACAA 8121

RESULT 9  
 ABL3246/c  
 ID ABL32466 standard; DNA; 15373 BP.  
 XX  
 AC ABL32466;  
 XX  
 DT 26-MAR-2002 (first entry)  
 DE Human immune system associated gene SEQ ID NO: 439.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antihaemic; cytosinetic; nootropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200928-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PE 02-JUL-2001; 2001WO-EP07537.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 DR WPI; 2002-130909/17.  
 XX  
 PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 XX  
 PS Claim 1; SEQ ID NO 439; 32pp + Sequence Listing; German.  
 XX  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 XX  
 SQ Sequence 15373 BP; 5235 A; 89 C; 2576 G; 7473 T; 0 other;

CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB57737-AB572072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 4025 BP; 1049 A; 1020 C; 933 G; 1023 T; 0 other;  
 QY Query Match 76.8%; Score 19.2; DB 23; Length 4025;  
 Best Local Similarity 87.5%; Pred. No. 2.1e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 DB 2 CATATATATCTACATCAAAACAA 25  
 3876 CATATATATCTACATCAAAACCA 3853  
 RESULT 6  
 ID AAS46781/C  
 AC AAS46781; standard; DNA; 4654 BP.  
 DT 18-DEC-2001 (first entry)  
 DE Tumour suppressor gene derived chemically modified sequence #506.  
 XX  
 KW Human: tumour suppressor gene; oncogene; antitumour; cytosstatic;  
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
 KW cytosine methylation; ds.  
 XX  
 OS Homo sapiens.  
 PN MO200168912-A2.  
 PD 20-SEP-2001.  
 PF 15-MAR-2001; 2001WO-EP02955.  
 PR 15-MAR-2000; 2000DE-1013847.  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 PA (EPIG-) EPIGENOMICS AG.  
 PI Olek A, Piepenbrock C, Berlin K;  
 DR WPI: 2001-602752/68.  
 XX  
 PT Fragments of chemically modified genes associated with tumour suppressor  
 PT genes and oncogenes, useful in designing primers and probes for  
 PT analysing diseases associated with cytosine methylation state e.g.  
 PT cancer.  
 XX  
 PS Claim 1; SEQ ID NO 506; 27pp; English.  
 CC The invention relates to a nucleic acid comprising a sequence of 18  
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
 CC bisulphite, of genes associated with tumour suppression and  
 CC oncogenes having a sequence taken from 536 (actually 533 since  
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a  
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
 CC form part of a set of probes for detecting the cytosine methylation state  
 CC and/or single nucleotide polymorphisms and also to be used in an  
 CC array for analysing diseases associated with CpG dinucleotides e.g.  
 CC cancers and tumours. The probes can also be used in a method for  
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
 CC and/or therapy of existing diseases or the predisposition to specific

CC diseases, by analysing cytosine methylations. The parameters may be  
 CC compared to another set of genetic and/or epigenetic parameters, the  
 CC differences serving as basis for diagnosis and/or prognosis events which  
 CC are disadvantageous to patients. The present sequence is one of the  
 CC 533 genomic sequences derived from tumour suppressor genes and  
 CC oncogenes. Sequences with even numbered Seq ID numbers are the  
 CC complementary sequence of the corresponding odd numbered sequence (e.g.  
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence  
 CC is missing).  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 4654 BP; 1313 A; 50 C; 955 G; 2336 T; 0 other;  
 QY Query Match 76.8%; Score 19.2; DB 22; Length 4654;  
 Best Local Similarity 87.5%; Pred. No. 2.1e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 DB 2 CATATATATCTACATCAAAACAA 25  
 874 CATATATATCTACATCAAAACCA 851  
 RESULT 7  
 ID ABL34223/C  
 AC ABL34223; standard; DNA; 4654 BP.  
 DT 26-MAR-2002 (first entry)  
 DE Human immune system associated gene SEQ ID NO: 2196.  
 XX  
 KW Human: immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antihaemic; cyostatic; noctropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 PN WO200200928-A2.  
 PD 03-JAN-2002.  
 PF 02-JUL-2001; 2001WO-EP07537.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 PA (EPIG-) EPIGENOMICS AG.  
 PI Olek A, Piepenbrock C, Berlin K;  
 DR WPI: 2002-130909/17.  
 XX  
 PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal,  
 PT cytosine methylation.  
 XX  
 PS Claim 1; SEQ ID NO 2196; 32pp + Sequence Listing; German.  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

PR 18-JUL-2000; 2000US-219007P.  
 PR 13-DEC-2000; 2000US-255281P.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Schlegel R, Endege WO, Monahan JE;  
 XX  
 DR WPI; 2001-662795/76.  
 XX  
 PT Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer; useful  
 PT for detecting presence of prostate cancer; stage of prostate cancer -  
 XX  
 PS Claim 1; Page 9866; 11750bp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement; (I) is useful for:  
 CC (a) assessing whether a patient is afflicted with prostate cancer;  
 CC (b) monitoring the progression of prostate cancer in a patient;  
 CC (c) assessing the efficacy of a test compound to inhibit prostate  
 CC cancer in a patient;  
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
 CC in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound;  
 CC (g) determining whether prostate cancer has metastasized in a patient;  
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
 CC patient;  
 CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
 XX  
 SQ Sequence 523 BP; 173 A; 108 C; 90 G; 151 T; 1 other:  
 XX  
 YQ Query Match 79.2%; Score 19.8; DB 23; Length 523;  
 Best Local Similarity 91.3%; Pred. No. 1.2e+02;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 XX  
 DB 3 ATATATCTACATTCATAACAA 25  
 411 AATATATCTACATTCATAACAA 389  
 XX  
 RESULT 4  
 AB067066/C  
 ID AB067066 standard; DNA; 5474 BP.  
 XX  
 AC AB067066;  
 XX  
 DT 28-AUG-2002 (first entry)  
 XX  
 DE Human angiogenesis associated polynucleotide SEQ ID NO 96.  
 XX  
 KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;  
 KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcers;  
 KW macular degeneration; inflammatory bowel disease; Crohn's disease;  
 KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiarteriosclerotic; ds.  
 XX  
 PA Homo sapiens.  
 XX  
 OS Homo sapiens.  
 XX  
 PI WO200246454-A2.  
 XX  
 PD 13-JUN-2002.  
 XX  
 PF 06-DEC-2001; 2001WO-EP14320.  
 XX  
 PR 06-DEC-2000; 2000DE-1061338.  
 XX  
 PS (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Schacht O;  
 XX  
 DR WPI; 2002-500450/53.

XX  
 PT New nucleic acid fragments from chemically treated  
 PT angiogenesis-associated genes; useful for determining methylation  
 PT status, e.g. in diagnosis or treatment of cancer -  
 XX  
 PS Claim 1; SEQ ID NO 96; 41pp + Sequence Listing; German.  
 XX  
 CC The invention relates to a nucleic acid (I) comprising a segment of 18  
 CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)  
 CC having sequences (AB066971-AB067178) or their complements. (I), also  
 CC related oligomers, are used to evaluate the methylation status and/or  
 CC single-nucleotide polymorphisms, in angiogenesis-related genes, for  
 CC diagnosis and treatment of eye diseases, proliferative retinopathy,  
 CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,  
 CC diabetic retinopathy, macular degeneration caused by neovascularisation,  
 CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and  
 CC Crohn's disease.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 5474 BP; 1440 A; 107 C; 1627 G; 2300 T; 0 other:  
 XX  
 YQ Query Match 79.2%; Score 19.8; DB 24; Length 5474;  
 Best Local Similarity 91.3%; Pred. No. 1.2e+02;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 XX  
 DB 2 CATATATCTACATTCATAACAA 24  
 1823 CATATATCTACATTCATAACAA 1801  
 XX  
 RESULT 5  
 ABLO5640/C  
 ID ABLO5640 standard; CDNA; 4025 BP.  
 XX  
 AC ABLO5640;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 11402.  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 11402.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PMD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 XX  
 P-PSDB; ABB61537.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 11402; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of

PT comprising the Casp8 gene that results in inactivation of the gene  
XX  
XX Example 3; Page 61; 107pp; English.  
XX  
CC Investigation of silencing of caspase-8 (CASP8) gene by methylation was  
CC analyzed using methylation-sensitive PCR analysis. Amplification of the  
CC 5' untranslated region (UTR) of the Casp8 gene was performed in reaction  
CC mixtures containing bisulfite treated DNA. Primers AAs51818-23 were  
CC designed to produce a 320 bp fragment in the upstream region of Casp8  
CC gene extending from nucleotides -221 to +541. Wild type primers were  
CC used to amplify the corresponding region of untreated genomic DNA.  
CC Controls without DNA were also performed. CASP8, a cysteine protease, is  
CC part of the death inducing signaling complex (DISC) associated with the  
CC Fas receptor. CASP8 is inactivated in cancers, and plays a role of a  
CC tumour suppressor gene. The CASP8 promoter region sequences, in  
CC particular Region 1 and Region 2, are crucial to the design and execution  
CC of the genomic methylation PCR analysis of CASP8 gene inactivation.  
CC Methylation PCR can be used to examine even minute amounts of patient  
CC material to demonstrate whether the CASP8 gene expresses an mRNA and  
CC protein product. The CASP8 gene has been localized to human chromosome  
CC 2q33-34. The methods are used to diagnose or prognose cancer. Cancer is  
CC treated by administering a vector that expresses a gene encoding  
CC functional CASP8 in cells. The cancer that is diagnosed or treated is a  
CC tumour in which a myc gene is amplified, such as a neuroblastoma.  
CC Aggressive neuroblastoma, juvenile neuroblastoma (preferred), small-cell  
CC lung carcinoma, non-small-cell lung carcinoma, colorectal carcinoma, or  
CC uterine cervical carcinoma can be diagnosed with the new method. A kit  
CC for screening for a compound that induces death-receptor-mediated  
CC apoptosis in cells containing an inactivated CASP8 gene is also  
CC provided.  
CC  
CC Sequence 25 BP; 12 A; 6 C; 0 G; 7 T; 0 other;  
SQ  
OY Query Match 100.0%; Score 25; DB 21; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 1 CCATATATCTACATCAAAACAA 25  
1 CCATATATCTACATCAAAACAA 25  
RESULT 2  
AAs46772/c  
ID AAs46772 standard; DNA: 17280 BP.  
XX  
AC AAs46772;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Tumour suppressor gene derived chemically modified sequence #496.  
XX  
XX  
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;  
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
XX cytosine methylation; ds.  
OS Homo sapiens.  
XX  
XX WO200168912-A2.  
PN 20-SEP-2001.  
XX  
XX 15-MAR-2001; 2001WO-EP02955.  
PF  
XX 15-MAR-2000; 2000DE-1013847.  
PR 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPig-) EPIGENOMICS AG.  
PA Olek A, Piepenbrock C, Berlin K;  
XX  
PI

XX  
DR WPI: 2001-602752/68.  
XX  
PT Fragments of chemically modified genes associated with tumour suppressor  
PT genes and oncogenes, useful in designing primers and probes for  
PT analysing diseases associated with cytosine methylation state e.g.  
PT cancer  
XX  
PS Claim 1; SEQ ID No 496; 27pp; English.  
XX  
CC The invention relates to a nucleic acid comprising a sequence of 18  
CC bases of a segment of chemically pretreated DNA (Cp DNA) e.g. with  
CC bisulphite, of genes associated with tumour suppression and  
CC oncogenes having a sequence taken from 536 (actually 533 since  
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a  
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
CC form part of a set of probes for detecting the cytosine methylation state  
CC and/or single nucleotide polymorphisms and also to be used in an  
CC array for analysing diseases associated with CpG dinucleotides e.g.  
CC cancers and tumours. The probes can also be used in a method for  
CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
CC and/or therapy of existing diseases or the predisposition to specific  
CC diseases, by analysing cytosine methylations. The parameters may be  
CC compared to another set of genetic and/or epigenetic parameters, the  
CC differences serving as basis for diagnosis and/or prognosis events which  
CC are disadvantageous to patients. The present sequence is one of the  
CC 533 genomic sequences derived from tumour suppressor genes and  
CC oncogenes. Sequences with even numbered Seq ID numbers are the  
CC complementary sequence of the corresponding odd numbered sequence (e.g.  
CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence  
CC is missing).  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
CC Sequence 17280 BP; 4730 A; 151 C; 3624 G; 8775 T; 0 other;  
SQ  
OY Query Match 80.8%; Score 20.2; DB 22; Length 17280;  
Best Local Similarity 88.0%; Pred. No. 88;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
DB 1 CCATATATCTACATCAAAACAA 25  
16705 CCAATATATCTACATTCAAAAAA 16681  
RESULT 3  
ABV50732/c  
ID ABV50732 standard; cDNA: 523 BP.  
XX  
AC ABV50732;  
XX  
DT 17-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA 50723.  
XX  
XX  
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200160860-A2.  
PN 23-AUG-2001.  
XX  
XX 20-FEB-2001; 2001WO-US05171.  
PF  
XX 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comogen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:03:20 ; Search time 24.9421 Seconds  
(without alignments)  
2257.229 Million cell updates/sec

Title: US-09-477-082-32  
Perfect score: 25  
Sequence: 1 ccatatatactacattcaaacaa 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	AAA51821	Antisense primer f
2	20.2	80.8	17280	AA546772	Tumour suppressor
3	19.8	79.2	523	ABV50732	Human prostate exp
4	19.8	79.2	5474	ABO67066	Human angiogenesis
5	19.2	76.8	4025	ABL05640	Drosophila melanog
6	19.2	76.8	4654	AA546781	Tumour suppressor
7	19.2	76.8	4654	ABL34223	Human immune syste
8	19.2	76.8	8845	AA546544	Tumour suppressor
9	19.2	76.8	15373	ABL32466	Human immune syste

C	10	18.8	75.2	5520	24	ABL33519	Human immune syste
C	11	18.8	75.2	6297	24	ABL32705	Human immune syste
C	12	18.8	75.2	9905	24	ABL32062	Human immune syste
C	13	18.6	74.4	221	22	AA161518	Soybean 240017 reg
C	14	18.6	74.4	361	22	AA191681	Human polynucleoti
C	15	18.6	74.4	379	22	AAK61674	Human immune/haema
C	16	18.6	74.4	798	24	ABK30312	Human G-protein-cc
C	17	18.6	74.4	1463	19	AAV61483	Human secreted pro
C	18	18.6	74.4	1463	24	ABO92010	Human polynucleoti
C	19	18.6	74.4	1567	22	ABA08735	Human secreted pro
C	20	18.6	74.4	1602	21	AAZ52465	Human clone 182123
C	21	18.6	74.4	4792	24	ABO54869	Human ovarian anti
C	22	18.6	74.4	5572	24	ABL54360	Chemically treated
C	23	18.6	74.4	5572	24	ABL33425	Human immune syste
C	24	18.6	74.4	6071	22	AA545373	Chemically pretrea
C	25	18.6	74.4	6071	24	ABK28210	DNA transcription
C	26	18.6	74.4	6960	22	AAK85579	Human immune/haema
C	27	18.6	74.4	7058	24	ABK40051	Human chemically p
C	28	18.6	74.4	10483	22	AAK80686	Human immune/haema
C	29	18.6	74.4	16236	24	ABL33023	Human immune syste
C	30	18.6	74.4	18011	21	AAK81914	Human immune syste
C	31	18.6	74.4	273254	22	AA161371	Chlamydia pneumoni
C	32	18.6	74.4	335913	22	AA161371	Soybean 240017 reg
C	33	18.6	74.4	335913	22	AA161372	Soybean 240017 reg
C	34	18.6	74.4	1230025	20	AAK01990	Nucleotide sequenc
C	35	18.4	73.6	10279	24	ABL32276	Chemically treated
C	36	18.4	73.6	10279	24	ABL33590	Human immune syste
C	37	18.4	73.6	10279	24	AAAD2327	Chemically treated
C	38	18.4	73.6	17294	24	ABL32987	Human immune syste
C	39	18.2	72.8	294	22	AA183128	Human polynucleoti
C	40	18.2	72.8	387	22	AA544850	Human cortig polyn
C	41	18.2	72.8	646	22	AAH87775	Peppermint plant o
C	42	18.2	72.8	672	23	ABL11307	Drosophila melanog
C	43	18.2	72.8	685	22	AA196618	Human neuroblastom
C	44	18.2	72.8	700	22	AAH92497	Human inflammatory
C	45	18.2	72.8	700	22	AAH92498	Human inflammatory

## ALIGNMENTS

RESULT 1  
ID AAA51821 standard; DNA; 25 BP.  
AC AAA51821:  
XX  
XX  
DT 31-OCT-2000 (first entry)  
XX  
XX  
DE Antisense primer for CASP8 unmethylated-specific disulfite treated DNA.  
XX  
XX  
KW CASP8; caspase-8; promoter; inactivation; methylation; cysteine protease;  
KW tumour suppressor; chromosome 2q33-34; neuroblastoma; cancer;  
KW death receptor; apoptosis; cytosolic; gene therapy; primer; ss.  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200039347-A1.  
XX  
XX  
PD 06-JUL-2000.  
XX  
XX  
PF 30-DEC-1999; 99WO-US31280.  
XX  
XX  
PR 31-DEC-1998; 98US-0114308.  
XX  
XX  
XX (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
XX  
XX  
XX Kidd VJ, Lahti JM, Teltz T;  
XX  
XX  
DR WPI; 2000-452423/39.  
XX  
XX  
PT Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or  
PT prognosing cancer, comprises detecting a modification of genomic DNA



Center:	Washington University Genome Sequencing Center
Center code:	WUSC
Web site:	http://genome.wustl.edu/gsc/index.shtml
Center project name:	H.NE0159F08
Summary Statistics	
Sequencing vector:	ML3; 100%
Sequencing vector:	plasmid; 0%
Chemistry:	Dye-terminator; 100% of reads
Chemistry:	Dye-terminator; Big dye; 0% of reads
Assembly program:	Pirap; version 0.990319
Consensus quality:	128904 bases at least Q40
Consensus quality:	133055 bases at least Q30
Consensus quality:	135453 bases at least Q20
Insert size:	122000; agarose-ftp
Insert size:	140332; sum-of-connigs
Quality coverage:	3.49 in Q20 bases; sum-of-connigs
NOTE:	This is a 'working draft' sequence. It currently consists of 28 connigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the connigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will still be preserved.
1	754: contig of 754 bp in length
755	854: gap of unknown length
855	1934: contig of 1080 bp in length
1935	4034: gap of unknown length
2035	4109: contig of 2075 bp in length
4110	4209: gap of unknown length
4210	5692: contig of 1483 bp in length
5693	5792: gap of unknown length
5793	7661: contig of 1469 bp in length
7262	7661: gap of unknown length
7362	9539: contig of 2178 bp in length
9540	9639: gap of unknown length
9640	11725: contig of 2086 bp in length
11726	11825: gap of unknown length
11826	14112: contig of 2887 bp in length
14113	14612: gap of unknown length
14613	17420: contig of 2808 bp in length
17421	17520: gap of unknown length
17521	20684: contig of 3164 bp in length
20685	20784: gap of unknown length
20785	23973: contig of 3189 bp in length
23974	24073: gap of unknown length
24074	26703: contig of 2630 bp in length
26704	26603: gap of unknown length
26804	29324: contig of 2521 bp in length
29325	29424: gap of unknown length
29425	32886: contig of 3462 bp in length
32887	32886: gap of unknown length
32887	36588: contig of 3402 bp in length

FEATURES	SOURCE
* 36389	36488: gap of unknown length
* 36489	41815: contig of 5527 bp in length
* 41815	41915: gap of unknown length
* 41915	47324: contig of 5409 bp in length
* 47324	47424: gap of unknown length
* 47425	51087: contig of 3663 bp in length
* 51088	51187: gap of unknown length
* 51188	56711: contig of 5524 bp in length
* 56712	56812: gap of unknown length
* 56812	61865: contig of 5054 bp in length
* 61865	61965: gap of unknown length
* 61965	68715: contig of 6750 bp in length
* 68716	68815: gap of unknown length
* 68816	78815: contig of 10000 bp in length
* 78816	78915: gap of unknown length
* 78915	86750: contig of 7835 bp in length
* 86751	86850: gap of unknown length
* 86851	97390: contig of 10540 bp in length
* 97391	97490: gap of unknown length
* 97491	106694: contig of 9204 bp in length
* 106695	106794: gap of unknown length
* 106795	116444: contig of 9650 bp in length
* 116445	116544: gap of unknown length
* 116545	128169: contig of 11625 bp in length
* 128170	128269: gap of unknown length
* 128270	140137: contig of 11868 bp in length
	Location/Qualifiers
	1..140137
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="12"
	/clone="RP11-159F8"
	1..754
misc_feature	/note="assembly_name:Contig10"
	855..1934
misc_feature	/note="assembly_name:Contig13"
	2035..4109
misc_feature	/note="assembly_name:Contig16"
	4210..5692
misc_feature	/note="assembly_name:Contig17"
	5793..7261
misc_feature	/note="assembly_name:Contig19"
	7362..9539
misc_feature	/note="assembly_name:Contig20"
	9640..11725
misc_feature	/note="assembly_name:Contig21"
	11826..14512
misc_feature	/note="assembly_name:Contig22"
	14613..17420
misc_feature	/note="assembly_name:Contig23"
	17521..20684
misc_feature	/note="assembly_name:Contig24"
	20785..23973
misc_feature	/note="assembly_name:Contig25"
	24074..26703
misc_feature	/note="assembly_name:Contig26
	vector_end:17
	vector_side:right"
misc_feature	26804..29324
misc_feature	/note="assembly_name:Contig27"
	29425..32886
misc_feature	/note="assembly_name:Contig28"
	32987..36388
misc_feature	/note="assembly_name:Contig29"
	36489..41815
misc_feature	/note="assembly_name:Contig30"
	41916..47324
misc_feature	/note="assembly_name:Contig31"
	47425..51087
misc_feature	/note="assembly_name:Contig33"
	51188..56711
misc_feature	/note="assembly_name:Contig33"
	56812..61865



```

misc-feature complement(9232..9417)
/misc-match: GSS: Em:AQ062104"
misc-feature 9243..9433
/misc-match: STS: Em:G31247"
misc-feature 9247..9358
/misc-match: STS: Em:G54157"
repeat-region 9433..9597
/note="AluJb repeat: matches 132..292 of consensus"
repeat-region 9631..9904
/note="AluJb repeat: matches 3..275 of consensus"
repeat-region 10052..10120
/note="MLT2D repeat: matches 490..553 of consensus"
repeat-region 10142..10183
/note="21 copies 2 mer ga 76% conserved"
repeat-region 10215..10651
/note="MLT2D repeat: matches 1..498 of consensus"
repeat-region 10671..10960
/note="AluSg repeat: matches 1..297 of consensus"
repeat-region 11084..11390
/note="AluSx repeat: matches 1..304 of consensus"
repeat-region 11722..11767
/note="L1P4 repeat: matches 5783..5828 of consensus"
misc-feature 11807..12289
/note="match: GSS: Em:AQ358081"
repeat-region 11872..12194
/note="AluSg repeat: matches 1..310 of consensus"
repeat-region 12355..12649
/note="AluSp repeat: matches 1..302 of consensus"
repeat-region 12652..12960
/note="AluY repeat: matches 1..308 of consensus"
repeat-region 13421..14475
/note="MER32C repeat: matches 1..1162 of consensus"
repeat-region 14476..14781
/note="AluSg repeat: matches 1..306 of consensus"
repeat-region 14782..14909
/note="MER52C repeat: matches 1162..1278 of consensus"
repeat-region 15276..15580

```

```

Query Match 80.8%; Score 20.2; DB 9; Length 25539;
Best Local Similarity 88.0%; Pred. No. 3.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 1 CCATATATATCTACATCAAAACAA 25
Db 18759 CCATATGATCTACATCAAAACAA 18783

```

```

RESULT 14
AL365396 114016 bp DNA linear PRI 23-AUG-2001
LOCUS Human DNA sequence from clone RP11-338N12 on chromosome 9, complete
DEFINITION
ACCESSION AL365396
VERSION AL365396.10 GI:15384808
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 114016)
AUTHORS Wray, P.
TITLE Direct Submission
JOURNAL CB10 15A, UK. E-mail enquiries: hunqurey@sanger.ac.uk

```

On Aug 31, 2001 this sequence version replaced gi:13990096. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30): an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMBASE; Information on the WORMBASE database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone configs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9> RP11-338N12 is from the library RPCT-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

#### FEATURES

```

source
1..114016
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-338N12"
/clone_lib="RPCT-11.2"
41280..41334
/note="Sequence from overlapping clone RP11-366H6 (AL449304). Assembly confirmed by restriction digest."
misc-feature 42423..42805
/note="Sequence from overlapping clone RP11-366H6 (AL449304). Assembly confirmed by restriction digest."
misc-feature 43815..43851
/note="Sequence from overlapping clone RP11-366H6 (AL449304). Assembly confirmed by restriction digest."
misc-feature 51123..51163
/note="Sequence from overlapping clone RP11-366H6 (AL449304). Assembly confirmed by restriction digest."
misc-feature 61961..62029
/note="Sequence from overlapping clone RP11-366H6 (AL449304). Assembly confirmed by restriction digest."
misc-feature 94609..94626
/note="Sequence from overlapping clone RP11-366H6 (AL449304). Assembly confirmed by restriction digest."
misc-feature 109428..109461
/note="Sequence from overlapping clone RP11-366H6 (AL449304). Assembly confirmed by restriction digest."
BASE COUNT 37485 a 19998 c 19402 g 37131 t
ORIGIN

```

```

Query Match 80.8%; Score 20.2; DB 9; Length 114016;
Best Local Similarity 88.0%; Pred. No. 2.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 1 CCATATATATCTACATCAAAACAA 25
Db 99446 CCATATGATCTACATCAAAACAA 99470

```

```

RESULT 15
AC027611 140137 bp DNA linear HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 12 clone RP11-155F8, WORKING DRAFT
DEFINITION
ACCESSION AC027611
VERSION AC027611.4 GI:8569832
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em', EMBL; Sw', SWISSPROT; Tr', TrEMBL; Mp', WormPep; Information on the WormPep database can be found at <http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/MGP/Chr13>

RP11-264M3 is from the library RPc11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6

**IMPORTANT:** This sequence is not the entire insert of clone RP11-264M3. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP11-74J13 is at 25440 in this sequence. The true right end of clone RP11-2P5 is at 100 in this sequence.

## FEATURES

```

source
1. .2539
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-264K3"
/clone_11b="RPci-11.1"
1. .174
repeat_region
/note="A1uXx repeat: matches 1. .174 of consensus"
175. .227
repeat_region
/note="MGTA repeat: matches 355. .380 of consensus"
228. .523
repeat_region
/note="A1uSg repeat: matches 1. .297 of consensus"
524. .562
repeat_region
/note="MGTA repeat: matches 380. .426 of consensus"
934. .1332
repeat_region
/note="MRR50 repeat: matches 10. .424 of consensus"
1472. .1582
repeat_region
/note="11P11 repeat: matches 6054. .6165 of consensus"
1583. .1885
repeat_region
/note="A1uSc repeat: matches 1. .303 of consensus"
1886. .2614
repeat_region
/note="11P11 repeat: matches 5315. .6054 of consensus"
2615. .2905
repeat_region
/note="A1uY repeat: matches 1. .292 of consensus"
2916. .3226
repeat_region
/note="A1uSx repeat: matches 1. .312 of consensus"
3227. .4486
repeat_region
/note="11P11 repeat: matches 4058. .5306 of consensus"
4487. .4660
repeat_region
/note="MRR50 repeat: matches 553. .734 of consensus"
5404. .5437
repeat_region
/note="17 copies 2 mer tt 88% conserved"
5454. .5626
repeat_region
/note="A1uSx repeat: matches 127. .299 of consensus"
5632. .5709
repeat_region
/note="39 copies 2 mer ta 85% conserved"
5712. .6012
repeat_region
/note="A1uY repeat: matches 1. .296 of consensus"
6014. .6148
repeat_region
/note="A1uSx repeat: matches 1. .133 of consensus"
6161. .6531
repeat_region

```

```
repeat_region /note="MER1D repeat: matches 1. .374 of consensus"
                    6533. .6754
/note="LTR9 repeat: matches 56. .282 of consensus"
repeat_region 7161. .7351
/note="MIR repeat: matches 6. .202 of consensus"
                    7772. .7895
/note="L1M repeat: matches 5687. .5818 of consensus"
misc_feature complement(7906. .8238)
                    /note="match: GSS: Em:B63021"
repeat_region 8010. .8288
/note="MER9 repeat: matches 1. .285 of consensus"
repeat_region 8388. .8586
/note="LTR9 repeat: matches 410. .609 of consensus"
repeat_region 8603. .8676
/note="Alus repeat: matches 11. .84 of consensus"
misc_feature complement(8807. .9429)
                    /note="match: GSS: Em:A788214"
repeat_region 8807. .8932
/note="AluB repeat: matches 1. .126 of consensus"
                    /complement(8816. .8944)
misc_feature /note="match: GSS: Em:AQ527616"
                    complement(8879. .9046)
misc_feature /note="match: STS: Em:G34094 Em:L00864"
                    complement(8908. .9106)
misc_feature /note="match: GSS: Em:AQ407117"
                    complement(8917. .9047)
misc_feature /note="match: STS: Em:G12329"
                    complement(8918. .9258)
misc_feature /note="match: GSS: Em:AQ267977"
                    complement(8925. .9093)
                    /note="match: GSS: Em:AQ781128"
misc_feature /note="match: GSS: Em:A2380835"
                    8927. .9093
                    /note="match: GSS: Em:A2380835"
                    8929. .9102
misc_feature /note="match: GSS: Em:AQ633135 Em:AQ89090"
                    complement(8932. .9068)
                    /note="match: STS: Em:HS442J2AS"
                    8933. .9018
                    /note="43 copies 2 mer ta 59% conserved"
                    8935. .9077
misc_feature /note="match: GSS: Em:A2254467"
                    complement(8935. .9064)
                    /note="match: GSS: Em:A2654134"
                    8941. .9081
                    /note="match: GSS: Em:B41348"
                    8961. .9078
                    /note="match: GSS: Em:AQ381413"
                    complement(8964. .9121)
                    /note="match: GSS: Em:A4707251"
                    complement(8965. .9121)
                    /note="match: GSS: Em:AQ077299"
                    8969. .9123
                    /note="match: GSS: Em:AQ589116"
                    complement(8970. .9121)
                    /note="match: STS: Em:G45335"
                    8977. .9092
                    /note="match: GSS: Em:A2773836"
                    complement(8982. .9225)
                    /note="match: GSS: Em:A2458983"
                    9001. .9281
                    /note="match: GSS: Em:AQ083235"
                    9020. .9121
                    /note="match: GSS: Em:A2396955"
                    9021. .9076
                    /note="28 copies 2 mer at 73% conserved"
                    complement(9100. .9417)
                    /note="match: GSS: Em:AQ782185"
                    9124. .9432
                    /note="Alus repeat: matches 1. .308 of consensus"
                    complement(9125. .9426)
                    /note="match: GSS: Em:AQ788085"
                    complement(9133. .9290)
                    /note="match: GSS: Em:AQ178343"
```

OY 4 TATATATCTACATTCAGAACAA 25  
 |||||  
 Db 136196 TATATATCTACATTCAGAACAA 136217

RESULT 11  
 AC122828  
 LOCUS 270987 bp DNA linear HTG 25-MAY-2002  
 DEFINITION Mus musculus chromosome UNK clone RP23-215B8, WORKING DRAFT  
 SEQUENCE 5 unordered pieces.

ACCESSION AC122828  
 VERSION AC122828.1 GI:21206486  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 270987)  
 McPherson, J.D. and Waterston, R.H.  
 TITLE The sequence of Mus musculus clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 270987)  
 McPherson, J.D. and Waterston, R.H.  
 AUTHORS Direct Submission  
 TITLE Submitted (25-MAY-2002) Genome Sequencing Center, 4444 Forest Park  
 JOURNAL Parkway, St. Louis, MO 63108, USA

COMMENT ----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 Contact: submissions@watson.wustl.edu  
 Project Information -----  
 Center project name: M.BA0215B08

----- Summary Statistics -----  
 Sequencing vector: M13; 0%  
 Sequencing vector: plasmid; 100%  
 Chemistry: Dye-primer ET; 0% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 268408 bases at least Q40  
 Consensus quality: 269116 bases at least Q20  
 Consensus quality: 269568 bases at least Q20  
 Insert size: 195000; agarose-fp  
 Quality coverage: 18.59 in Q20 bases; agarose-fp  
 Quality coverage: 10.33 in Q20 bases; sum-of-coverage

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1472: contig of 1472 bp in length  
 \* 1473 1572: gap of unknown length  
 \* 1573 2810: contig of 1238 bp in length  
 \* 2811 2910: gap of unknown length  
 \* 2911 47630: contig of 44720 bp in length  
 \* 47631 47730: gap of unknown length  
 \* 47731 107746: contig of 60016 bp in length  
 \* 107747 107846: gap of unknown length  
 \* 107847 270987: contig of 163141 bp in length.

FEATURES  
 source  
 1. 270987  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="UNK"  
 /clone="RP23-215B8"  
 misc-feature 1. 1472

misc-feature /note="assembly\_name:Contig5"  
 1573. 2810  
 /note="assembly\_name:Contig6"  
 misc-feature 2911. 47630  
 /note="assembly\_name:Contig7"  
 misc-feature 47731. 107746  
 /note="assembly\_name:Contig8"  
 misc-feature 107847. 270987  
 /note="assembly\_name:Contig9"

BASE COUNT 80382 a 56456 c 54330 g 79404 t 415 others  
 ORIGIN

Query Match 81.6%; Score 20.4; DB 2; Length 270987;  
 Best Local Similarity 95.5%; Pred. No. 1.9e+02;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 TATATATCTACATTCAGAACAA 25  
 |||||  
 Db 262929 TATATATCTACATTCAGAACAA 262950

RESULT 12  
 AX251530/c 17280 bp DNA linear PAT 05-OCT-2001  
 LOCUS  
 DEFINITION Sequence 498 from Patent WO0168912.  
 ACCESSION AX251530  
 VERSION AX251530.1 GI:15984953  
 KEYWORDS  
 SOURCE synthetic construct.  
 ORGANISM  
 synthetic construct.  
 artificial sequences.  
 1 (bases 1 to 17280)  
 REFERENCE Olek, A., Piepenbrock, C. and Berlin, K.  
 AUTHORS Diagnosis of diseases associated with tumor suppressor genes and  
 TITLE Oncogenes  
 JOURNAL Patent: WO 0168912-A 498 20-SEP-2001;  
 EpiGenomics AG (DE)  
 FEATURES  
 source  
 1. 17280  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 4730 a 151 c 3624 g 8775 t  
 ORIGIN

Query Match 80.8%; Score 20.2; DB 6; Length 17280;  
 Best Local Similarity 88.0%; Pred. No. 4.1e+02;  
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 CCAATATATCTACATTCAGAACAA 25  
 |||||  
 Db 16705 CCAATATATCTACATTCAGAACAA 16681

RESULT 13  
 AL161775 25539 bp DNA linear PRI 25-APR-2001  
 LOCUS Human DNA sequence from clone RP11-264M3 on chromosome 13, complete  
 DEFINITION sequence.  
 ACCESSION AL161775  
 VERSION AL161775.20 GI:13872274  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 25539)  
 Lovell, J.  
 REFERENCE Direct Submission  
 AUTHORS Submitted (25-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 JOURNAL CB10 ISA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 On Apr 27, 2001 this sequence version replaced gi:13396429.

KEYWORDS HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE house mouse.  
ORGANISM Mus musculus.

REFERENCE Mammalia: Eutharia; Chordata: Craniata; Vertebrata: Euteleostomi;  
TITLE 1 (bases 1 to 226726)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
JOURNAL Mus musculus, clone RP23-28P10  
REFERENCE Unpublished  
ATTNORS 2 (bases 1 to 226726)

Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,  
Anderson,S., Barina,N., Bastien,C., Bloom,T., Boguslavsky,L.,  
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,  
Lander,E., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
Maclean,C., Macdonald,P., Major,J., Marguis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Menus,L.,  
Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trifilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission  
JOURNAL Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome  
REFERENCE Research, 320 Charles Street, Cambridge, MA 02141, USA  
AUTHORS 3 (bases 1 to 226726)

Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,  
Anderson,S., Barina,N., Bastien,C., Bloom,T., Boguslavsky,L.,  
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,  
Lander,E., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
Maclean,C., Macdonald,P., Major,J., Marguis,N.,  
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J.,  
Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trifilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission  
JOURNAL Submitted (26-JUN-2002) Whitehead Institute/MIT Center for Genome  
COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 26, 2002 this sequence version replaced gi:21536040.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information  
Center Project name: L20899  
Center clone name: 28\_P\_10

----- Summary Statistics  
Sequencing Vector: Plasmid; N/A; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 224168 bases at least Q40  
Consensus quality: 225210 bases at least Q40  
Consensus quality: 225616 bases at least Q20  
Insert size: 240000; agarose-1p  
Insert size: 225926; sum-of-coverage  
Quality coverage: 7.9 in Q20 bases; agarose-1p  
Quality coverage: 8.3 in Q20 bases; sum-of-coverage

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 1496: contig of 1496 bp in length  
\* 1497 1596: gap of 100 bp  
\* 1597 1853: contig of 257 bp in length  
\* 1854 1953: gap of 100 bp  
\* 1954 5276: contig of 3323 bp in length  
\* 5277 5376: gap of 100 bp  
\* 5377 9084: contig of 3708 bp in length  
\* 9085 9184: gap of 100 bp  
\* 9185 25659: contig of 16475 bp in length  
\* 25660 25759: gap of 100 bp  
\* 25760 48397: contig of 22638 bp in length  
\* 48398 48497: gap of 100 bp  
\* 48498 144109: contig of 95612 bp in length  
\* 144110 144209: gap of 100 bp  
\* 144210 200142: contig of 55933 bp in length  
\* 200143 200242: gap of 100 bp  
\* 200243 226726: contig of 26484 bp in length.

## FEATURES

source  
1..226726  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-28P10"  
/clone.lib="RP23-28P10"  
Location/Qualifiers  
1..1496  
/note="assembly-fragment"  
clone\_end:Sp6  
vector\_side:left  
1597..1853  
/note="assembly-fragment"  
1954..5276  
/note="assembly-fragment"  
5377..9084  
/note="assembly-fragment"  
9185..25659  
/note="assembly-fragment"  
25760..48397  
/note="assembly-fragment"  
48498..144109  
/note="assembly-fragment"  
144210..200142  
/note="assembly-fragment"  
200243..226726  
/note="assembly-fragment"  
clone\_end:17  
vector\_side:right  
BASE COUNT 65243 a 45772 c 45704 g 69207 t 800 others  
ORIGIN

Query Match 81.6%; Score 20.4; DB 2; Length 226726;  
Best Local Similarity 93.5%; Pred. No. 2e+02; Indels 0; Gaps 0;  
Matches 21; Conservative 0; Mismatches 1;